



Heriot-Watt University
Research Gateway

Non-asymptotic bounds for sampling algorithms without log-concavity

Citation for published version:

Majka, MB, Mijatović, A & Szpruch, L 2020, 'Non-asymptotic bounds for sampling algorithms without log-concavity', *Annals of Applied Probability*, vol. 30, no. 4, pp. 1534-1581. <https://doi.org/10.1214/19-AAP1535>

Digital Object Identifier (DOI):

[10.1214/19-AAP1535](https://doi.org/10.1214/19-AAP1535)

Link:

[Link to publication record in Heriot-Watt Research Portal](#)

Document Version:

Peer reviewed version

Published In:

Annals of Applied Probability

General rights

Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact open.access@hw.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Non-asymptotic bounds for sampling algorithms without log-concavity

Mateusz B. Majka^{1,2}, Aleksandar Mijatović^{1,2}, and Łukasz Szpruch^{1,3}

¹The Alan Turing Institute, London

²Department of Statistics, University of Warwick

³School of Mathematics, University of Edinburgh

Abstract

Discrete time analogues of ergodic stochastic differential equations (SDEs) are one of the most popular and flexible tools for sampling high-dimensional probability measures. Non-asymptotic analysis in the L^2 Wasserstein distance of sampling algorithms based on Euler discretisations of SDEs has been recently developed by several authors for log-concave probability distributions. In this work we replace the log-concavity assumption with a log-concavity at infinity condition. We provide novel L^2 convergence rates for Euler schemes, expressed explicitly in terms of problem parameters. From there we derive non-asymptotic bounds on the distance between the laws induced by Euler schemes and the invariant laws of SDEs, both for schemes with standard and with randomised (inaccurate) drifts. We also obtain bounds for the hierarchy of discretisation, which enables us to deploy a multi-level Monte Carlo estimator. Our proof relies on a novel construction of a coupling for the Markov chains that can be used to control both the L^1 and L^2 Wasserstein distances simultaneously. Finally, we provide a weak convergence analysis that covers both the standard and the randomised (inaccurate) drift case. In particular, we reveal that the variance of the randomised drift does not influence the rate of weak convergence of the Euler scheme to the SDE.

1 Introduction

Our primary aim is to study non-asymptotic properties of Markov chains that typically arise as approximations of ergodic solutions of stochastic differential equations on \mathbb{R}^d . The simplest example is a process $(X_k)_{k=0}^\infty$ defined as

$$\begin{cases} X_{k+1} = X_k + b(X_k)h + \sqrt{h}\xi_{k+1}, & k \geq 0, \\ X_0 \sim \mu_0, \end{cases} \quad (1.1)$$

where $h > 0$ is the discretisation parameter and $(\xi_k)_{k=1}^\infty$ are i.i.d. random variables with the standard normal distribution $N(0, I)$. Here $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$, the space of square integrable probability measures on \mathbb{R}^d , and $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a drift function. We use the notation $\mathcal{L}(X_k) := \text{Law}(X_k)$ and our main goal is to quantify convergence of $\mathcal{L}(X_k)$ using the L^p -Wasserstein distance with $p \in \{1, 2\}$, defined for probability measures $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ as

$$W_p(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \pi(dx dy) \right)^{1/p}, \quad (1.2)$$

where $\Pi(\mu, \nu)$ denotes the family of all couplings between μ and ν , i.e., all measures on $\mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d)$ such that $\pi(B \times \mathbb{R}^d) = \mu(B)$ and $\pi(\mathbb{R}^d \times B) = \nu(B)$ for every $B \in \mathcal{B}(\mathbb{R}^d)$.

Here $|\cdot| = \sqrt{\langle \cdot, \cdot \rangle}$ is the Euclidean distance on \mathbb{R}^d . We will also work with a special class of L^1 -Wasserstein (pseudo) distances denoted by W_f in which the Euclidean distance $|x - y|$ is replaced by $f(|x - y|)$ for some increasing function $f : [0, \infty) \rightarrow [0, \infty)$. Namely, we put $W_f(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} f(|x - y|) \pi(dx dy)$. We remark that Wasserstein distances are typically preferred metrics when quantifying the quality of sampling methods, see [ACB17, DK19, GDVM16].

Convergence in Wasserstein distances is typically investigated under the contractivity condition on the drift, i.e., under the assumption that there exists a constant $K > 0$ such that

$$\langle x - y, b(x) - b(y) \rangle \leq -K|x - y|^2 \quad \text{for all } x, y \in \mathbb{R}^d. \quad (1.3)$$

If $b(x) = -\nabla U(x)$ for a function $U \in C^2(\mathbb{R}^d)$, this condition corresponds to strong convexity of U , whereas a probability measure μ such that $\mu(dx) \propto \exp(-U(x))dx$ is then called log-concave. Convergence analysis for several sampling algorithms under condition (1.3) and the Lipschitz continuity of the drift has been recently performed in the L^2 -Wasserstein distance in papers such as [DK19, BFFN19, DM16, CFM⁺18].

In this work instead of (1.3) we work with the following assumptions.

Assumptions 1.1 (Contractivity at infinity). Function $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ satisfies the following conditions:

i) Lipschitz condition: there is a constant $L > 0$ such that

$$|b(x) - b(y)| \leq L|x - y| \quad \text{for all } x, y \in \mathbb{R}^d. \quad (1.4)$$

ii) Contractivity at infinity condition: there exist constants $K, \mathcal{R} > 0$ such that

$$\langle x - y, b(x) - b(y) \rangle \leq -K|x - y|^2 \quad \text{for all } x, y \in \mathbb{R}^d \text{ with } |x - y| > \mathcal{R}. \quad (1.5)$$

This enables us to cover a much wider class of SDEs, including e.g. equations with drifts given by double-well potentials (see the example in Section 2.1). We will show in Section 2.1 that tools typically used in the global contractivity setting to study convergence in Wasserstein distances, such as the synchronous coupling and Talagrand's inequality, do not necessarily work under Assumptions 1.1. This forces us to look for an alternative approach.

Our method is based on the idea of controlling the standard Wasserstein distances W_1 and W_2 by specially constructed (pseudo) distances W_f based on functions $f : [0, \infty) \rightarrow [0, \infty)$ that are concave on a compact interval and strictly convex at infinity. In order to briefly introduce our approach, let us focus on a single step X_h^x of an Euler scheme started at $x \in \mathbb{R}^d$, i.e., $X_h^x = x + b(x)h + \sqrt{h}\xi$, where $\xi \sim N(0, I)$. If we now choose $y \in \mathbb{R}^d$, $y \neq x$ and consider $Y_h^y = y + b(y)h + \sqrt{h}\eta$ with an arbitrarily chosen $\eta \sim N(0, I)$, then $\mathcal{L}(Y_h^y) = \mathcal{L}(X_h^y)$ and straight from the definition of the Wasserstein distance W_f we see that $W_f(\mathcal{L}(X_h^x), \mathcal{L}(X_h^y)) \leq \mathbb{E}f(|X_h^x - Y_h^y|)$, since the joint law of (X_h^x, Y_h^y) is a coupling of $\mathcal{L}(X_h^x)$ and $\mathcal{L}(X_h^y)$. Hence in order to obtain sharp upper bounds on $W_f(\mathcal{L}(X_h^x), \mathcal{L}(X_h^y))$, one needs to be able to find pairs (X_h^x, Y_h^y) that make $\mathbb{E}f(|X_h^x - Y_h^y|)$ as small as possible by choosing the joint distribution of (ξ, η) in an appropriate way. However, in the present paper we are interested not only in quantifying distances between laws of Euler schemes started at different points, but also in distances between laws of Euler schemes and their perturbed versions. Namely, let \tilde{Y} be an arbitrarily chosen random variable. Under Assumptions 1.1 we are able to prove the following inequality (see Theorem 2.5 and the comments after its proof) for all sufficiently small $h > 0$ with constants $c, C > 0$.

$$\mathbb{E}f(|X_h^x - \tilde{Y}|) \leq (1 - ch)f(|x - y|) + C(1 + |x - y|)(\mathbb{E}|\tilde{Y} - Y_h^y| + \mathbb{E}|\tilde{Y} - Y_h^y|^2). \quad (1.6)$$

The idea behind this result is that if \tilde{Y} is chosen as a perturbation of Y_h^y such that we can control the distance between \tilde{Y} and Y_h^y in the L^1 and L^2 norms, then we can also control the distance between \tilde{Y} and X_h^x , via an auxiliary pseudo distance function f that dominates all L^p distances for $p \in [1, 2]$. The exact form of the function f and all the constants will be given in Theorem 2.5. The important fact is that the function f is chosen in such a way that there exist constants a, A such that $r < af(r)$ and $r^2 < Af(r)$ for $r \in [0, \infty)$ (see Lemma 2.3) and hence (1.6) yields bounds for W_1 and W_2 distances.

Our inequality (1.6) is related to other perturbation results for Wasserstein distances, see the discussion in Section 2.3 of [EM19] and papers such as [RS18, PS14, HZ17]. As we explain in the sequel, examples of processes that we consider in this paper as perturbed versions of Euler schemes include an SDE (Section 2.3), an Euler scheme with a randomised drift (Section 2.4) and an Euler scheme with a different discretisation level (Section 2.5).

The Markov chain given by (1.1) arises as a discretisation of a diffusion process given by

$$dY_t = b(Y_t) dt + dW_t, \quad (1.7)$$

where $(W_t)_{t \geq 0}$ is a standard Brownian motion. Assumptions 1.1 guarantee that the solution does not blow up on $[0, \infty)$, see Chapter 3 in [Kha12]. It can be shown that the corresponding semigroup is Feller and consequently the Krylov-Bogolyubov theorem yields the existence of a unique invariant measure π . We remark that the asymptotic results on the discretisation of (1.7) in the context of approximating invariant measures are rather well understood. We refer to [Tal90, LP02, Pan08, PP12] for a thorough overview on that topic.

Working with the global contractivity (1.3) assumption and the Euler discretisation (1.1) with either constant or variable time-step, several authors in a series of papers [Dal17, DM17, DM16, CB18] obtained precise bounds on $W_2(\mathcal{L}(X_k), \pi)$ in terms of dimension and problem parameters. These bounds have been then improved in [DK19]. Here we obtain precise convergence rates in the L^1 and L^2 -Wasserstein distances working only with Assumptions 1.1. In particular, our bound is reminiscent of the bounds in [DK19, DM16] for the constant step size algorithms. Indeed we show the following result.

Theorem 1.2. *Suppose that Assumptions 1.1 are satisfied. Then there exist a function f and constants $h_0, \hat{c}_1, \hat{c}_2, C_1, C_2, \tilde{c}_1, \tilde{c}_2 > 0$ such that for all $h \in (0, h_0)$*

$$W_2(\mathcal{L}(X_k), \pi) \leq C_2(1 - \hat{c}_2 h)^{k/2} (W_f(\mathcal{L}(X_0), \pi))^{1/2} + \tilde{c}_2 h^{1/4} \quad (1.8)$$

and

$$W_1(\mathcal{L}(X_k), \pi) \leq C_1(1 - \hat{c}_1 h)^k W_f(\mathcal{L}(X_0), \pi) + \tilde{c}_1 h^{1/2}. \quad (1.9)$$

The precise values of all the constants and the formula defining the function f in Theorem 1.2 will be given in Section 2.3.

Our next observation is that the perturbation inequality (1.6) sheds light onto stochastic gradient algorithms or Langevin models with inaccurate/randomised gradients. Examples of such processes have been used in [WT11] in the context of sampling and studied in [DK19, RRT17, TTV16, VZT16, NDH⁺17]. We remark that randomisation is a successful technique that is known to improve convergence properties for problems with non-smooth coefficients [KW17, PM14, JN09]. We define a function $\bar{b} : \mathbb{R}^d \times \mathbb{R}^n \rightarrow \mathbb{R}^d$. Consider an \mathbb{R}^n -valued random variable U such that $\mathbb{E}[\bar{b}(x, U)] = b(x)$ for all $x \in \mathbb{R}^d$. Let $(\xi_k)_{k=1}^\infty$, as before, be i.i.d. with $\xi_k \sim N(0, I)$ and take i.i.d. copies $(U_k)_{k=0}^\infty$ of U that are independent of $(\xi_k)_{k=1}^\infty$. We define the following Markov chain

$$\begin{cases} \bar{X}_{k+1} = \bar{X}_k + \bar{b}(\bar{X}_k, U_k)h + \sqrt{h}\xi_{k+1}, & k \geq 0, \\ \bar{X}_0 \sim \mu_0. \end{cases} \quad (1.10)$$

Note that for each k , the random variable \bar{X}_k is independent of U_k and that $\mathbb{E}[\bar{b}(\bar{X}_k, U_k)|\bar{X}_k] = b(\bar{X}_k)$. For each k we consider the conditional variance $\mathbb{V}[\bar{b}(\bar{X}_k, U_k)|\bar{X}_k] := \mathbb{E}[|\bar{b}(\bar{X}_k, U_k) - b(\bar{X}_k)|^2|\bar{X}_k]$, which as we shall see plays a key role in our analysis in Section 2.4. We need to impose the following assumption.

Assumption 1.3 (Variance of inaccurate drift). There exist constants $\sigma, \alpha > 0$ such that for any $x \in \mathbb{R}^d$, any $h > 0$ and any random variable U such that $\mathbb{E}[\bar{b}(x, U)] = b(x)$, we have

$$\mathbb{E} |\bar{b}(x, U) - b(x)|^2 \leq \sigma^2(1 + |x|^2)h^\alpha. \quad (1.11)$$

Note that the dependence on h of the right hand side of (1.11) is related to the choice of the estimator \bar{b} . On the other hand, the constants σ and α do not depend on h . We will discuss how to verify Assumption 1.3 in the case where the drift is estimated by subsampling in Example 2.15.

We can study properties of $(\bar{X}_k)_{k=0}^\infty$ by treating it as a perturbed version of $(X_k)_{k=0}^\infty$ given by (1.1). This allows us to study convergence of $\mathcal{L}(\bar{X}_k)$ to π . Indeed, using the fact that W_2 satisfies the triangle inequality, we have $W_2(\mathcal{L}(\bar{X}_k), \pi) \leq W_2(\mathcal{L}(\bar{X}_k), \mathcal{L}(X_k)) + W_2(\mathcal{L}(X_k), \pi)$ and the bound on $W_2(\mathcal{L}(\bar{X}_k), \mathcal{L}(X_k))$ will follow from inequality (1.6). Hence we obtain the following result.

Theorem 1.4. Let Assumptions 1.1 and 1.3 hold. Then there exist constants $h_0, \bar{C}_1, \bar{C}_2 > 0$ such that for all $h \in (0, h_0)$

$$W_2(\mathcal{L}(\bar{X}_k), \pi) \leq W_2(\mathcal{L}(X_k), \pi) + \bar{C}_2 h^{\alpha/4} \quad (1.12)$$

and

$$W_1(\mathcal{L}(\bar{X}_k), \pi) \leq W_1(\mathcal{L}(X_k), \pi) + \bar{C}_1 h^{\alpha/2}. \quad (1.13)$$

The precise values of the constants and the proof of Theorem 1.4 can be found in Section 2.4. Note that our bounds in Theorem 1.4 are of similar form as the bounds obtained in Theorem 4 in [DK19] in a setting corresponding to the global contractivity assumption (1.3), i.e., in [DK19] the distance $W_2(\mathcal{L}(\bar{X}_k), \pi)$ is also bounded by $W_2(\mathcal{L}(X_k), \pi)$ plus an additional error term coming from the use of an inaccurate drift. However, the error term in W_2 in the global contractivity case in [DK19] is obtained under an assumption similar to our (1.11) with $\alpha = 0$ and is of order $\sigma h^{1/2}$, whereas our error term is of order $\sigma h^{\alpha/4}$ in W_2 and $\sigma^2 h^{\alpha/2}$ in W_1 for $\alpha \geq 0$ (dependence on σ of the constants \bar{C}_1 and \bar{C}_2 can be easily traced in Section 2.4). We believe that the worse order of the constants in our case is a necessary consequence of the much more general contractivity at infinity assumption, however, it remains an open question whether our constants are actually optimal. To our knowledge, the only related result without assuming global contractivity of the drift was obtained in [RRT17], see Proposition 10 therein, by using functional inequalities. However, the estimates in [RRT17] are not uniform in time.

Note that the bounds we obtain in Theorem 1.4 depend on the variance of the estimator \bar{b} of the drift b via h^α appearing in (1.11). This is in contrast with the following weak convergence result, which at least in the context of randomised Euler schemes seems to be new.

Theorem 1.5. Let Assumption 1.3 hold. Let $g \in C^\infty(\mathbb{R}^d, \mathbb{R})$ with polynomial growth and assume that $b \in C^3$ has bounded derivatives and that there are constants $M_1, M_2 > 0$ such that for all $x \in \mathbb{R}^d$ we have

$$\langle x, b(x) \rangle \leq M_2 - M_1 |x|^2. \quad (1.14)$$

Furthermore, assume that for any $p \geq 1$ there is a constant $C_b^p > 0$ such that for any $x \in \mathbb{R}^d$ and for any random variable U with $\mathbb{E}[\bar{b}(x, U)] = b(x)$ we have

$$\mathbb{E} |\bar{b}(x, U)|^p \leq C_b^p (1 + |x|^p). \quad (1.15)$$

Then there exists a constant $c_w > 0$ independent of h such that

$$\sup_{k \in \mathbb{N}} |\mathbb{E}[g(Y_{kh})] - \mathbb{E}[g(\bar{X}_k)]| \leq c_w h.$$

Note that condition (1.14) is implied by (1.5), see also Lemma 2.11. Moreover, condition (1.15) for all $p \geq 1$ is a relatively weak assumption that is implied e.g. by the Lipschitz continuity of $\bar{b}(\cdot, U)$ (cf. (1.16)) and a moment bound on $\bar{b}(0, U)$ for all U as in (1.15). The proof of Theorem 1.5, unlike all the other results in this paper, does not rely on the perturbation inequality (1.6) and instead uses PDE methods and estimates from [Tal90]. The proof can be found in Section 4. The above result, in addition to being interesting on its own, will be also applied in our analysis of Multi-level Monte Carlo (MLMC) in Section 2.5.

Non-asymptotic results on the L^2 -Wasserstein distance pave the way to the analysis of the MLMC method, which is a variance reduction technique introduced in [Hei01], [Keb05] and [Gil08]. Let π be the invariant measure of (1.7) and let $g \in C^\infty(\mathbb{R}^d, \mathbb{R})$ be Lipschitz. Our framework allows us to consider multi-level type estimators of $\int_{\mathbb{R}^d} g(x) \pi(dx)$, involving Euler schemes with inaccurate drifts, which have been treated before only in [GMS⁺19]. However, the authors of [GMS⁺19] considered exclusively the globally contractive case, whereas in the present paper we employ the coupling method in order to deal with the more general contractivity at infinity setting. The multi-level estimator will be introduced in detail in Section 2.5. For our results on MLMC with inaccurate drift we need to impose additional assumptions.

Assumptions 1.6 (MLMC with inaccurate drift). The function $\bar{b} : \mathbb{R}^d \times \mathbb{R}^n \rightarrow \mathbb{R}^d$ satisfies the following conditions:

- i) Lipschitz condition: There is a constant $\bar{L} > 0$ such that for all $x, y \in \mathbb{R}^d$ and all random variables U such that $\mathbb{E}[\bar{b}(x, U)] = b(x)$ and $\mathbb{E}[\bar{b}(y, U)] = b(y)$ we have

$$|\bar{b}(x, U) - \bar{b}(y, U)| \leq \bar{L}|x - y| \text{ a.s.} \quad (1.16)$$

- ii) Contractivity at infinity condition: There exist constants $\bar{K}, \bar{R} > 0$ such that for all $x, y \in \mathbb{R}^d$ with $|x - y| \geq \bar{R}$ and for all random variables U such that $\mathbb{E}[\bar{b}(x, U)] = b(x)$ and $\mathbb{E}[\bar{b}(y, U)] = b(y)$ we have

$$\langle x - y, \bar{b}(x, U) - \bar{b}(y, U) \rangle \leq -\bar{K}|x - y|^2 \text{ a.s.} \quad (1.17)$$

We have the following result.

Theorem 1.7. *Let all the assumptions of Theorem 1.5 hold. Additionally, suppose that Assumptions 1.6 are satisfied and that g is Lipschitz. Then for the estimation of $\int_{\mathbb{R}^d} g(x) \pi(dx)$ by MLMC with inaccurate drift we have computational complexity $\mathcal{O}(\epsilon^{-2-(1-\min\{1, \alpha\})/2} \log \epsilon)$ with α given in (1.11), as opposed to complexity $\mathcal{O}(\epsilon^{-3} \log \epsilon)$ in the standard Monte Carlo approach.*

The proof of Theorem 1.7 and a detailed description of our MLMC estimator can be found in Section 2.5. Note that our results apply also to MLMC for Euler schemes with non-randomised drifts. There $\alpha = \infty$ and we obtain a gain in complexity from $\mathcal{O}(\epsilon^{-3} \log \epsilon)$ to $\mathcal{O}(\epsilon^{-5/2} \log \epsilon)$ under Assumptions 1.1.

Remark 1.8. While we were working on the present paper, Fang and Giles in [FG19] independently obtained results on MLMC for SDEs with non-globally contractive drifts using a change of measure argument. Their approach, contrary to ours, does not aim at obtaining bounds on the variance that are uniform in time. While their variance grows linearly with time, they have better strong convergence rates and hence in the accurate drift case overall complexity of their algorithm is better than here. However, our framework is more flexible and allows us to cover also Euler schemes with inaccurate drifts.

Remark 1.9. While completing this work, the paper [CCA⁺18] appeared, where results analogous to our Theorem 1.2 were obtained under assumptions on the drift similar to ours, although only in the L^1 -Wasserstein distance. More precisely, the focus in [CCA⁺18] is on determining the smallest number of iterations of the Euler scheme required to approach the invariant measure of the SDE (1.7) in the W_1 distance with precision ε . In our setting, we can infer from our Theorem 1.2 that this number of iterations is of order $\mathcal{O}(d/\varepsilon^2)$, which is the same as in [CCA⁺18]. To see this, note that the constant \tilde{c}_1 in (1.9) is of order $\mathcal{O}(\sqrt{d})$, cf. (2.29) and (2.35), and hence a similar analysis as in the proof of Theorem 2.1 in [CCA⁺18] gives the required result. Interestingly, for an analogous problem in the L^2 Wasserstein distance, our estimates give us a required number of iterations of order $\mathcal{O}(d^2/\varepsilon^4)$, since the constant \tilde{c}_2 in (1.8) is also of order $\mathcal{O}(\sqrt{d})$, but is multiplied by $h^{1/4}$ instead of $h^{1/2}$. However, we do not know whether this result is sharp or just an artifact of our proof. We are not aware of any comparable estimates in the W_2 distance in the non-convex setting in the literature. Furthermore, we remark that determining optimality of such bounds is usually a non-trivial task, as it requires finding lower bounds for the Wasserstein distances, see e.g. Example 4 in [Ebe16] for a related discussion in the diffusion case.

2 Contractivity of Euler schemes and applications

Before we present our results in detail, let us briefly discuss why the classical methods may fail if we do not assume global log-concavity.

2.1 Motivation

Synchronous coupling It is well known that under the global convexity assumption (1.3) on the drift, we can show that Euler schemes are contractive in the L^2 distance just by applying the synchronous coupling. Namely, if we have

$$X' := x + hb(x) + \sqrt{h}Z =: \hat{x} + \sqrt{h}Z, \quad (2.1)$$

where $Z \sim N(0, I)$ and we define

$$Y' := y + hb(y) + \sqrt{h}Z =: \hat{y} + \sqrt{h}Z,$$

then for all sufficiently small h we obtain

$$\mathbb{E}|X' - Y'|^2 = |\hat{x} - \hat{y}|^2 \leq (1 - ch)|x - y|^2$$

for some constant $c \in (0, 1)$. However, under the dissipativity at infinity assumption (1.5) the synchronous coupling is no longer sufficient. As a one-dimensional example, consider the function

$$U(x) := x^2(g_n(x))^2 + a^2 - 2axg_n(x)$$

for some parameters $a > 0$ and $n \geq 1$, where

$$g_n(x) := \begin{cases} -n & \text{if } x \in (-\infty, -n) \\ x & \text{if } x \in [-n, n] \\ n & \text{if } x \in (n, \infty) \end{cases}.$$

The function U is constructed by truncating the function $x \mapsto (x^2 - a)^2$ so that U' is Lipschitz. We consider the drift

$$b(x) = -U'(x).$$

It is easy to check that b satisfies the dissipativity at infinity assumption. Indeed, when $|x|, |y| \leq n$ then we have

$$\langle x - y, b(x) - b(y) \rangle = -4(x - y)^2(x^2 + xy + y^2 - a) \quad (2.2)$$

and we see that there exist constants $R_0, C > 0$ such that we have

$$\langle x - y, b(x) - b(y) \rangle \leq -C(x - y)^2$$

when $|x - y| > R_0$. Similarly, when $|x|, |y| \geq n$, we get

$$\langle x - y, b(x) - b(y) \rangle = -2n(x - y)^2 + 2an \left(\frac{x}{|x|} - \frac{y}{|y|} \right) (x - y)$$

and again if the distance $|x - y|$ is large enough, then the desired inequality holds. When $y > n > x$ then the dissipativity condition

$$\left(-4x^3 + 4ax + 2ny - 2an \frac{y}{|y|} \right) (x - y) \leq -C(x - y)^2$$

is equivalent to

$$-4x^3 + (4a + C)x + (2n - C)y - 2an \frac{y}{|y|} \geq 0,$$

which we can make sure is satisfied by choosing the parameters in an appropriate way, and the other cases follow by symmetry. However, from (2.2) we also see that there exist constants $R_1, R_2 > 0$ such that when $|x|, |y| \leq R_1$ and $|x - y| \leq R_2$, we have

$$\langle x - y, b(x) - b(y) \rangle \geq C_1(x - y)^2$$

for some constant $C_1 > 0$. Hence we see that if we use the synchronous coupling for our Euler scheme, we end up having

$$\begin{aligned} \mathbb{E}|X' - Y'|^2 &= |\hat{x} - \hat{y}|^2 = |x - y|^2 + h^2|b(x) - b(y)|^2 + 2h(b(x) - b(y))(x - y) \\ &\geq |x - y|^2 + 2hC_1|x - y|^2 \end{aligned}$$

and thus we cannot have a contraction.

Talagrand inequality An alternative approach to proving contractivity of Euler schemes in the L^2 -Wasserstein distance in the global convexity setting relies on the fact that the Gaussian measure μ with covariance matrix Σ satisfies the Talagrand inequality with the constant being the largest eigenvalue $\lambda_{\max}(\Sigma)$, i.e., for any measure ν absolutely continuous with respect to μ we have

$$W_2(\nu, \mu)^2 \leq 2\lambda_{\max}(\Sigma)H(\nu|\mu) = 2\lambda_{\max}(\Sigma) \int \log \frac{d\nu}{d\mu} d\nu, \quad (2.3)$$

see e.g. [Tal96] or page 2726 in [DGW04] and the references therein. Namely, taking $\nu = N(x + hb(x), hI)$ and $\mu = N(y + hb(y), hI)$ we have

$$\begin{aligned} H(\nu|\mu) &= \int \log \exp \left(\frac{-(z - (x + hb(x)))^2}{2h} \right) \exp \left(\frac{(z - (y + hb(y)))^2}{2h} \right) \nu(dz) \\ &= \frac{1}{2h} \int (-z^2 + 2z(x + hb(x)) - (x + hb(x))^2) + (z^2 - 2z(y + hb(y)) + (y + hb(y))^2) \nu(dz) \\ &= \frac{1}{2h} ((x + hb(x))^2 - 2(x + hb(x))(y + hb(y)) + (y + hb(y))^2) \\ &= \frac{1}{2h} (x + hb(x) - y - hb(y))^2 \end{aligned}$$

and hence the right hand side of (2.3) is equal to $(\hat{x} - \hat{y})^2$. In the global convexity setting this can be bounded from above by $(1 - ch)|x - y|^2$ for all sufficiently small $h > 0$. However, in the non-convex setting we can use the same example as the one presented above, where $(\hat{x} - \hat{y})^2$ is bounded from below by $|x - y|^2 + 2hC_1|x - y|^2$ with some $C_1 > 0$. This shows that the approach via Talagrand's inequality fails as well.

2.2 Coupling and Wasserstein (pseudo) distances

The random vector (X_h^x, Y_h^y) in (1.6) is an example of a *coupling* of random variables. Constructing different random objects with the same distributions is a widely applied tool in probability theory, see e.g. [Lin92, Tho00] for general overview and [EM19, NP19] and the references therein for other examples of applications of couplings to sampling algorithms.

In a series of works [Ebe11, Ebe16], Eberle used the reflection coupling for diffusions to prove W_1 contractivity for the SDE (1.7) under Assumptions 1.1. More precisely, he proved that there exists a constant $\lambda > 0$, expressed explicitly in terms of problem parameters, such that

$$W_1(\mathcal{L}(Y_t^\mu), \mathcal{L}(Y_t^\nu)) \leq e^{-\lambda t} W_1(\mu, \nu), \quad (2.4)$$

where Y_t^μ denotes the solution to (1.7) with $Y_0 \sim \mu$. By taking $\nu = \pi$ in (2.4), i.e., by initialising the process at the invariant measure π , one immediately obtains geometric convergence rate of the law of the process $(Y_t)_{t \geq 0}$ to its stationary measure.

In the present paper we extend the ideas from [EM19], where results analogous to the ones from [Ebe16] have been obtained directly on the level of the Markov chain (1.1) in W_1 , by employing a suitably chosen coupling and constructing a special Kantorovich (L^1 -Wasserstein) distance, see Section 2.4 therein. Here we introduce a novel coupling construction and a new Wasserstein-type pseudo-distance, which allows us to obtain L^2 bounds on (1.1) as well as on its perturbed version, and hence to analyse convergence of several sampling algorithms presented below.

One of the benefits of working with the L^2 -Wasserstein distance is that through the Kantorovich duality theory (see e.g. Theorem 5.10 in [Vil09]) it significantly extends the class of functions g for which we can obtain explicit convergence rates of functionals $g(X_k)$ of Euler schemes (1.1). When working with the L^1 Wasserstein distance, duality theory gives access only to Lipschitz functions, whereas in our setup we can also consider e.g. functions of quadratic growth. Moreover, L^2 bounds are necessary for our analysis of the variance of multi-level Monte Carlo estimators in Section 2.5.

We fix $h > 0$ and we consider one step of the Euler scheme for the equation (1.7), started at a point $x \in \mathbb{R}^d$, i.e., we have a random variable X' given by

$$X' = \hat{x} + \sqrt{h}Z, \quad (2.5)$$

where

$$\hat{x} = x + hb(x)$$

is the deterministic movement due to the drift b and $Z \sim N(0, I)$, where I is the $d \times d$ identity matrix. Hence we have $X' \sim N(x + hb(x), hI)$. For a given point $y \neq x$, we want to construct a new random variable Y' that will have the distribution $N(y + hb(y), hI)$. We define

$$\hat{y} = y + hb(y)$$

and

$$\hat{r} = |\hat{x} - \hat{y}|.$$

We want to define a coupling of the random movement $(\hat{x}, \hat{y}) \mapsto (X', Y')$. Note that in [EM19] bounds in L^1 were obtained by choosing

$$Y' = \begin{cases} X', & \text{if } \zeta \leq (\varphi_{\hat{y},hI}(X') \wedge \varphi_{\hat{x},hI}(X')) / \varphi_{\hat{x},hI}(X'), \\ \hat{y} + R_{\hat{x},\hat{y}}\sqrt{h}Z, & \text{if } \zeta > (\varphi_{\hat{y},hI}(X') \wedge \varphi_{\hat{x},hI}(X')) / \varphi_{\hat{x},hI}(X'). \end{cases} \quad (2.6)$$

Here ζ is a uniformly distributed random variable on $[0, 1]$ independent of Z , $R_{x,y} = I - 2(x - y)(x - y)^T / |x - y|^2$ is the reflection operator with respect to the hyperplane spanned by $(x - y) / |x - y|$ (note that if the dimension $d = 1$ then $R_{x,y}u = -u$ for any $u \in \mathbb{R}$) and $\varphi_{z,A}(u)$ is the density of $N(z, A)$ for $z \in \mathbb{R}^d$ and $A \in \mathbb{R}^{d \times d}$. We call (2.6) the mirror coupling.

Note that this type of coupling is related to the one used in the optimal transport theory in [McC99], where it was shown to be the optimal coupling for all concave costs in the one-dimensional case, for a class of probability measures that includes Gaussian measures (but is in fact much broader). This makes (2.6) the right choice of coupling for obtaining optimal bounds in concave Wasserstein distances. However, it needs to be modified to work for convex distances such as W_2 .

In the present paper we use a combination of the synchronous coupling and the coupling given by (2.6), defined in the following way.

We choose two parameters $H > 0$ and $m > 0$. If $\hat{r} > H$, we choose the synchronous coupling, i.e., we set $Y' = \hat{y} + \sqrt{h}Z$. If $\hat{r} \leq H$, we compare the size of the jump (i.e., the size of $\sqrt{h}Z$) with the parameter m . If $|\sqrt{h}Z| > m$, we again define Y' synchronously. Otherwise, we use the variable ζ to apply the mirror coupling defined by (2.6) to the Gaussians truncated by m .

To be precise, we have

$$X' = \hat{x} + \sqrt{h}Z$$

$$Y' = \begin{cases} X', & \text{if } \zeta \leq \left(\varphi_{\hat{y},hI}^m(X') \wedge \varphi_{\hat{x},hI}^m(X') \right) / \varphi_{\hat{x},hI}^m(X') \text{ and } |\sqrt{h}Z| < m \text{ and } \hat{r} \leq H, \\ \hat{y} + R_{\hat{x},\hat{y}}\sqrt{h}Z, & \text{if } \zeta > \left(\varphi_{\hat{y},hI}^m(X') \wedge \varphi_{\hat{x},hI}^m(X') \right) / \varphi_{\hat{x},hI}^m(X') \text{ and } |\sqrt{h}Z| < m \text{ and } \hat{r} \leq H, \\ \hat{y} + \sqrt{h}Z, & \text{if } |\sqrt{h}Z| \geq m \text{ or } \hat{r} > H. \end{cases} \quad (2.7)$$

Here $\varphi_{z,A}^m(u) := \mathbf{1}_{\{|u-z| \leq m\}} \varphi_{z,A}(u)$. Note that we only need to evaluate $\varphi_{\hat{x},hI}^m(X')$ when $|\sqrt{h}Z| < m$ (or, equivalently, when $|X' - \hat{x}| < m$) and hence we always have $\varphi_{\hat{x},hI}^m(X') \neq 0$, which implies that all the quantities in (2.7) are well-defined. It is easy to prove that (2.7) is indeed a coupling of $N(x + hb(x), hI)$ and $N(y + hb(y), hI)$, see Theorem 6.1 in the Appendix.

Note that Y' defined above can be thought of as a function of the initial points $x, y \in \mathbb{R}^d$ and the random variable Z . It also depends on the truncation parameters m and H . In the sequel we will use the shorthand notation

$$\psi_{m,H}(x, y, Z) := Y' \quad (2.8)$$

for the random variable Y' defined by (2.7).

Our goal is to expand on the methods introduced in [EM19], where, under the same assumptions on the drift b as in the present paper, the coupling (X', Y') given by (2.6) was used in order to obtain estimates of the form

$$\mathbb{E}f(|X' - Y'|) \leq (1 - ch)f(|x - y|)$$

for a specially constructed increasing, concave function f . Since the function f in [EM19] is comparable both from above and from below with a rescaled identity function, this implies

bounds for the first moment $\mathbb{E}|X' - Y'|$, as well as contractivity of the laws of the Euler scheme in a Kantorovich (L^1 Wasserstein) distance. The latter is a strong property with multiple important consequences, see e.g. [JO10] or the discussion in [EM19]. Here we use a modified coupling and a different construction of a distance function $f : [0, \infty) \rightarrow [0, \infty)$, which in our case is concave up to some $r_2 > 0$ and convex afterwards. This allows us to get upper bounds for the second moment $\mathbb{E}|X' - Y'|^2$, cf. Lemma 2.3.

Related work in the diffusion case has been done in [LW16], where the authors adapted the L^1 bounds from [Ebe16] and obtained L^p bounds for $p \geq 1$ under similar assumptions as in [Ebe16]. Here, however, we use a different, more direct construction of the auxiliary distance function f . We also introduce a novel coupling construction, which is specific to the discrete time case.

Before we formulate our main result, let us introduce some notation. We define positive constants

$$c_0 := 4 \min \left(\int_0^{1/2} u^2 (1 - e^{u-1/2}) \varphi_{0,1}(u) du, (1 - e^{-1}) \int_0^{1/2} u^3 \varphi_{0,1}(u) du \right),$$

where $\varphi_{0,1}$ is the density of $N(0, 1)$, and

$$h_0 := \min \left(\frac{K}{L^2}, \frac{4}{K}, \frac{1}{2L}, \frac{2c_0 \ln \frac{3}{2}}{27L^2 \mathcal{R}^2}, \frac{\mathcal{R}^2}{4}, \frac{c_0^2 (\ln 2)^2}{144L^2 \mathcal{R}^2} \right), \quad (2.9)$$

where the constants L, K and $\mathcal{R} > 0$ are all specified in Assumptions 1.1. Then we put $r_1 := (1 + h_0 L) \mathcal{R}$ and $r_2 := r_1 + \sqrt{h_0}$. Finally, we choose the parameter

$$a := \frac{6Lr_1}{c_0} \quad (2.10)$$

and we construct our function $f : [0, \infty) \rightarrow [0, \infty)$ by setting

$$f(r) := \begin{cases} \frac{1}{a}(1 - e^{-ar}) & \text{if } r \leq r_2 \\ \frac{1}{a}(1 - e^{-ar_2}) + \frac{1}{2r_2}e^{-ar_2}(r^2 - r_2^2) & \text{if } r > r_2 \end{cases}. \quad (2.11)$$

Then we have the following result.

Theorem 2.1. *Let Assumptions 1.1 hold. For the coupling (X', Y') given by (2.7) with parameters $m = \frac{\sqrt{h_0}}{2}$ and $H = r_1$ and the function f given by (2.11), we have*

$$\mathbb{E}f(|X' - Y'|) \leq (1 - ch)f(|x - y|) \quad (2.12)$$

for all $h \in (0, h_0)$, where

$$c = \min \left(e^{-ar_2} \frac{K}{4}, \frac{\frac{1}{2}e^{-ar_2}r_2}{\frac{1}{a}(1 - e^{-ar_2})} \frac{K}{4}, \frac{9L^2r_1^2}{2c_0} e^{-\frac{6Lr_1^2}{c_0}}, \frac{3Lr_1}{16\sqrt{h_0}} \right). \quad (2.13)$$

The proof will be presented in Section 3. As an immediate consequence, we obtain bounds in the W_f Wasserstein (pseudo) distance.

Corollary 2.2. *Let $p(x, A)$ for $x \in \mathbb{R}^d$ and $A \in \mathcal{B}(\mathbb{R}^d)$ denote the transition kernel of X' given by (2.5) and let $\mu p(dy) := \int \mu(dx) p(x, dy)$ for any probability measure μ on \mathbb{R}^d . Under the assumptions of Theorem 2.1, we have*

$$W_f(\mu p, \nu p) \leq (1 - ch)W_f(\mu, \nu)$$

for all $h \in (0, h_0)$ and for all $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$.

The above result is a straightforward consequence of (2.12) and the definition of W_f , see also the proof of Lemma 2.1 in [EM19].

An important feature of the function f given by (2.11) is that it is comparable from below with the identity and the square functions. This allows us to obtain both L^1 and L^2 bounds as an immediate consequence of Theorem 2.1.

Lemma 2.3. *We have*

$$r^2 \leq Af(r) \text{ and } r \leq e^{ar_2} f(r) \quad (2.14)$$

for all $r \in [0, \infty)$, where

$$A := \max \left(\frac{ar_2^2}{1 - e^{-ar_2}}, 2r_2 e^{ar_2} \right). \quad (2.15)$$

Proof. The second bound in (2.14) is obvious since $f'(r) \geq e^{-ar_2}$ for all $r \geq 0$. Now let us try to look for a constant $A > 0$ such that

$$r_2^2 \leq Af(r_2) = A \frac{1}{a} (1 - e^{-ar_2}). \quad (2.16)$$

If this holds, then for any $r \leq r_2$ we have $r^2 \leq Af(r)$. Hence we need $A \geq ar_2^2 / (1 - e^{-ar_2})$. Moreover, we need $r^2 = r_2^2 - r_2^2 + r^2 \leq A \frac{1}{a} (1 - e^{-ar_2}) + A \frac{1}{2r_2} e^{-ar_2} (r^2 - r_2^2)$ for any $r \geq r_2$, hence, using (2.16), we see that we need $r^2 - r_2^2 \leq A \frac{1}{2r_2} e^{-ar_2} (r^2 - r_2^2)$, which implies $A \geq 2r_2 e^{ar_2}$. \square

Corollary 2.4. *Under the assumptions of Theorem 2.1, we have*

$$W_2(\mu p, \nu p) \leq A(1 - ch)W_f(\mu, \nu) \quad \text{and} \quad W_1(\mu p, \nu p) \leq e^{ar_2}(1 - ch)W_f(\mu, \nu)$$

for all $h \in (0, h_0)$ and for all $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, where A is given by (2.15) and all the other constants are as in Theorem 2.1.

An important consequence of Theorem 2.1 that turns out to be crucial for our applications, is the following inequality.

Theorem 2.5. *Let (X', Y') be the coupling given by (2.7) with parameters $m = \frac{\sqrt{h_0}}{2}$ and $H = r_1$ and let f be the function given by (2.11), where all the constants are as specified in Theorem 2.1. Let \tilde{X} be a random variable. Then*

$$\begin{aligned} \mathbb{E}f(|Y' - \tilde{X}|) &\leq (1 - ch)f(|y - x|) + \mathbb{E} \left[\frac{1}{r_2} e^{-ar_2} |X' - \tilde{X}|^2 \right] \\ &\quad + \mathbb{E} \left[\left(1 + \frac{1}{r_2} e^{-ar_2} \left(|y - x|(1 + hL) + \sqrt{h_0} \right) \right) |X' - \tilde{X}| \right]. \end{aligned} \quad (2.17)$$

Proof. We have

$$\mathbb{E}f(|Y' - \tilde{X}|) - f(|y - x|) = \mathbb{E}f(|Y' - \tilde{X}|) - \mathbb{E}f(|Y' - X'|) + \mathbb{E}f(|Y' - X'|) - f(|y - x|).$$

By (2.12) we know that $\mathbb{E}f(|Y' - X'|) - f(|y - x|) \leq -chf(|y - x|)$ and hence it is sufficient to focus on the expression $\mathbb{E}f(|Y' - \tilde{X}|) - \mathbb{E}f(|Y' - X'|)$. Since f is increasing, we have

$$\mathbb{E}f(|Y' - \tilde{X}|) - \mathbb{E}f(|Y' - X'|) \leq \mathbb{E}f(|Y' - X'| + |X' - \tilde{X}|) - \mathbb{E}f(|Y' - X'|). \quad (2.18)$$

We can now apply the Taylor formula to see that the right hand side of (2.18) is equal to $\mathbb{E}f'(\zeta)|X' - \tilde{X}|$, where $\zeta \in (|Y' - X'|, |Y' - X'| + |X' - \tilde{X}|)$. From the definition of f we get $f'(\zeta) \leq \frac{1}{r_2} e^{-ar_2} |\zeta| + e^{-a|\zeta|} \leq \frac{1}{r_2} e^{-ar_2} |\zeta| + 1$ and hence

$$\mathbb{E}f'(\zeta)|X' - \tilde{X}| \leq \mathbb{E} \left[\left(\frac{1}{r_2} e^{-ar_2} (|Y' - X'| + |X' - \tilde{X}|) + 1 \right) |X' - \tilde{X}| \right].$$

Now observe that due to our construction of the coupling (X', Y') we have $|Y' - X'| \leq |\hat{y} - \hat{x}| + 2m$. This is because an increase in the distance between the two coupled processes can happen in the random step $(\hat{x}, \hat{y}) \mapsto (X', Y')$ only due to reflection. However, reflection happens only if the normal random variable Z takes a sufficiently small value and hence that increase is bounded by $2m$. Moreover, due to the Lipschitz condition (1.4) of the drift we have $|\hat{y} - \hat{x}| \leq |y - x|(1 + hL)$. Recall that in order for (2.12) to hold, we choose $m = \sqrt{h_0}/2$. Combining all our estimates together, we arrive at (2.17). \square

From the result above we see that if we choose \tilde{X} in such a way that we have control on the first and the second moment of the distance between \tilde{X} and X' , then we can get a good bound on the distance between Y' and \tilde{X} . Hence our result is useful for random variables \tilde{X} that are small perturbations of X' . Note that the role of X' and Y' in the proof of Theorem 2.5 is symmetric, hence we immediately obtain (1.6). The reason for our choice of the form of inequality (2.17) will become apparent in the proofs in Sections 2.3, 2.4 and 2.5.

We can also obtain a related perturbation result based on Theorem 2.12 from [EM19], which is a result similar to our Theorem 2.1 but with a globally concave function f . This allows us to get simpler formulas in cases where we only need L^1 bounds. Let us define $q = 7c_0^{-1}L\mathcal{R}$ and

$$f_1(r) := \begin{cases} \frac{1}{q}(1 - e^{-qr}) & \text{if } r \leq r_1^1 \\ \frac{1}{q}(1 - e^{-qr_1^1}) + e^{-qr_1^1}(r - r_1^1) & \text{if } r > r_1^1 \end{cases}, \quad (2.19)$$

where $r_1^1 := (1 + h_0^1 L)\mathcal{R}$ with h_0^1 specified in Theorem 2.6. Then we have the following result.

Theorem 2.6. *Let Assumptions 1.1 hold. Then for the coupling (X', Y') given by (2.7) with parameters $m = \infty$ and $H = \infty$ and the function f_1 given by (2.19), we have*

$$\mathbb{E}f_1(|X' - Y'|) \leq (1 - c_1 h)f_1(|x - y|) \quad (2.20)$$

for all $h \in (0, h_0^1)$, where

$$\begin{aligned} c_1 &= \min \left(\frac{K}{2}, \frac{245}{24c_0} L^2 \mathcal{R}^2 \right) \exp \left(-\frac{49}{6c_0} L \mathcal{R}^2 \right) \quad \text{and} \\ h_0^1 &= \frac{1}{L} \min \left(\frac{1}{6}, \frac{K}{L}, \frac{1}{3} L \mathcal{R}^2, \frac{c_0^2}{970} \frac{1}{L \mathcal{R}^2} \right). \end{aligned} \quad (2.21)$$

Note that the exact statement of Theorem 2.12 in [EM19] is slightly different than Theorem 2.6 above, since the metric in [EM19] depends on h . This can be easily modified by replacing the constant $\Lambda = \Lambda(h)$ defined by (2.59) in [EM19] with L and $r_1 = r_1(h)$ defined by (2.65) therein with $r_1^1 := (1 + h_0^1 L)\mathcal{R}$. Then the proof in [EM19] easily carries over to our setting and we obtain Theorem 2.6. We leave the details to the interested reader. More importantly, from inequality (2.20) we can easily derive another perturbation result.

Theorem 2.7. *Let Assumptions 1.1 hold and let \tilde{X} be a random variable. Then for all $h \in (0, h_0^1)$ we have*

$$\mathbb{E}f_1(|Y' - \tilde{X}|) \leq (1 - c_1 h)f_1(|y - x|) + \mathbb{E}|X' - \tilde{X}|.$$

Proof. Note that the function f_1 defined in (2.19) is concave (and hence subadditive), increasing and its derivative is such that $f_1'(x) \in [e^{-qr_1^1}, 1]$ for all $x \in \mathbb{R}_+$. Thus we get

$$\begin{aligned} \mathbb{E}f_1(|Y' - \tilde{X}|) &\leq \mathbb{E}f_1(|Y' - X'| + |X' - \tilde{X}|) \leq \mathbb{E}f_1(|Y' - X'|) + \mathbb{E}f_1(|X' - \tilde{X}|) \\ &\leq (1 - c_1 h)f_1(|y - x|) + \mathbb{E}|X' - \tilde{X}|. \end{aligned}$$

\square

Even though most results in the present paper are based on the contractivity and the perturbation theorems presented above for Euler schemes $X' = x + hb(x) + \sqrt{h}Z$, it may sometimes be useful to be able to couple processes with inaccurate drift and obtain respective counterparts of Theorems 2.1 and 2.5. In the sequel we will need such results only in Section 2.5 where we treat Multi-level Monte Carlo algorithms for the inaccurate drift case. To this end, consider

$$\bar{X}' = x + h\bar{b}(x, U) + \sqrt{h}Z \quad (2.22)$$

and define $\hat{x} = x + h\bar{b}(x, U)$ and $\hat{y} = y + h\bar{b}(y, U)$, where the random variable U is independent of Z and such that $\mathbb{E}\bar{b}(x, U) = b(x)$ for all $x \in \mathbb{R}^d$. Hence, we can still use the prescription (2.7) to define a coupling (\bar{X}', \bar{Y}') of two copies of (2.22) started from arbitrary points $x, y \in \mathbb{R}^d$ by substituting \hat{x} and \hat{y} defined for b that appear in (2.7) with our new \hat{x} and \hat{y} defined for \bar{b} . In other words, using similar notation as in (2.8), we have a new transformation

$$\bar{Y}' = \bar{\psi}_{m,H}(x, y, U, Z) \quad (2.23)$$

that we can use to couple two copies of the process given by (2.22). Since the statement about equality of marginal laws in the coupling given by (2.7) is actually a statement about the Gaussian steps $\hat{x} \mapsto X'$ and $\hat{y} \mapsto Y'$ and we assume independence of U and Z , in order to prove that the pair (\bar{X}', \bar{Y}') obtained via (2.23) is indeed a coupling of two copies of (2.22), we can apply the reasoning from Theorem 6.1 by replacing the expectation there with the conditional expectation with respect to U . For the coupling (\bar{X}', \bar{Y}') obtained this way, we have the following result.

Theorem 2.8. *Suppose Assumptions 1.6 are satisfied with constants $\bar{K} = K$, $\bar{R} = \mathcal{R}$ and $\bar{L} = L > 0$. Then for the coupling (\bar{X}', \bar{Y}') of two copies of (2.22) defined via (2.23) we have*

$$\mathbb{E} [f(|\bar{X}' - \bar{Y}'|) | U] \leq (1 - ch)f(|x - y|) \text{ a.s.}$$

for all $h \in (0, h_0)$, with the same constants h_0, c, m, H and the same function f as in Theorem 2.1. We also have the perturbation inequality (2.17) in the unchanged form, with (X', Y') replaced by (\bar{X}', \bar{Y}') .

The proof of Theorem 2.8 will be presented in Section 3. Note that Theorem 2.8 actually implies Theorems 2.1 and 2.5, since the Euler scheme (2.5) can be easily interpreted as a special case of the scheme (2.22) with inaccurate drift. Hence we need to prove only Theorem 2.8.

2.3 Unadjusted Langevin algorithm

In this section we will explain how to apply Theorem 2.5 in order to obtain bounds on the distance between the invariant measure of the solution $(Y_t)_{t \geq 0}$ of the SDE

$$dY_t = b(Y_t)dt + dW_t, \quad (2.24)$$

where $(W_t)_{t \geq 0}$ is a d -dimensional Brownian motion, and the laws of the Markov chain $(X_k)_{k=0}^\infty$ given by

$$X_{k+1} = X_k + hb(X_k) + \sqrt{h}\xi_{k+1}, \quad (2.25)$$

for all $k \geq 0$, where $h > 0$ is fixed and $(\xi)_{k=1}^\infty$ are i.i.d. random variables with the standard normal distribution.

If $b(x) = -\frac{1}{2}\nabla U(x)$ for a function $U \in C^2(\mathbb{R}^d)$, then the equation (2.24) is called the (overdamped) Langevin SDE. Its invariant measure is given by

$$\pi(dz) := \frac{\exp(-U(z))dz}{\int_{\mathbb{R}^d} \exp(-U(x))dx},$$

see e.g. [DM17] or Example 1 in [Ebe16]. The method of asymptotic sampling from π by using the Markov chain defined by (2.25) is called the Unadjusted Langevin Algorithm.

In order to proceed, let us first observe that

$$Y_{(k+1)h} = Y_{kh} + \int_{kh}^{(k+1)h} b(Y_s)ds + (W_{(k+1)h} - W_{kh}) \quad (2.26)$$

for all $k \geq 0$. We have the following result.

Theorem 2.9. *Under the assumptions of Theorem 2.1, for any random variables X_0, Y_0 , any $k \geq 1$ and for any $h \in (0, h_0 \wedge \frac{K}{4L^2})$ we have*

$$W_2(\mathcal{L}(X_k), \mathcal{L}(Y_{kh})) \leq \left(A(1 - ch)^k \mathbb{E}f(|X_0 - Y_0|) \right)^{1/2} + \left(\frac{AC_{ult}}{c} \right)^{1/2} h^{1/4},$$

whereas for any $h \in (0, h_0^1)$ we have

$$W_1(\mathcal{L}(X_k), \mathcal{L}(Y_{kh})) \leq e^{qr_1} (1 - c_1 h)^k \mathbb{E}f_1(|X_0 - Y_0|) + e^{qr_1} \frac{\sqrt{C_{dif}}}{c_1} h^{1/2}, \quad (2.27)$$

where c is given by (2.13), c_1 and h_0^1 are given by (2.21), A is given by (2.15), C_{dif} is given by (2.35) and C_{ult} is defined in (2.36). In particular, if $Y_0 \sim \pi$, then for any $k \geq 1$ and any $h \in (0, h_0 \wedge \frac{K}{4L^2})$ we have

$$W_2(\mathcal{L}(X_k), \pi) \leq \left(A(1 - ch)^k \mathbb{E}f(|X_0 - Y_0|) \right)^{1/2} + \left(\frac{AC_{ult}}{c} \right)^{1/2} h^{1/4}, \quad (2.28)$$

whereas for any $h \in (0, h_0^1)$ we have

$$W_1(\mathcal{L}(X_k), \pi) \leq e^{qr_1} (1 - c_1 h)^k \mathbb{E}f_1(|X_0 - Y_0|) + e^{qr_1} \frac{\sqrt{C_{dif}}}{c_1} h^{1/2}. \quad (2.29)$$

Remark 2.10. Note that in Theorem 2.9 we apply our Theorem 2.1 only to obtain L^2 bounds, while L^1 bounds are obtained using Theorem 2.6 taken from [EM19]. We could still apply Theorem 2.1 in the L^1 case and obtain

$$W_1(\mathcal{L}(X_k), \mathcal{L}(Y_{kh})) \leq e^{ar_2} (1 - ch)^k \mathbb{E}f(|X_0 - Y_0|) + e^{ar_2} \frac{C_{ult}}{c} h^{1/2}$$

instead of (2.27). However, (2.27) gives us better dependence of the constants on the dimension d , since C_{ult} is of order $\mathcal{O}(d)$, whereas $\sqrt{C_{dif}}$ is of order $\mathcal{O}(\sqrt{d})$.

Proof of Theorem 2.9. We define

$$\sqrt{h}Z_{k+1} := W_{(k+1)h} - W_{kh}$$

and consider an auxiliary process $(S_k)_{k=1}^\infty$ given by

$$S_{k+1} := Y_{kh} + hb(Y_{kh}) + \sqrt{h}Z_{k+1}. \quad (2.30)$$

We can think of $(S_k)_{k=1}^\infty$ as a process which at each step k is moved to the position of $(Y_{kh})_{k=1}^\infty$ and then evolves in the same way as (2.25), with the random noise given by an increment of the

Brownian motion $(W_t)_{t \geq 0}$. Now consider a new process $(G_k)_{k=0}^\infty$ coupled to $(S_k)_{k=1}^\infty$ according to the prescription (2.7). Using the notation introduced in (2.8), we set

$$G_{k+1} := \psi_{m,H}(Y_{kh}, G_k, Z_{k+1}). \quad (2.31)$$

We can alternatively describe $(G_k)_{k=0}^\infty$ as

$$G_{k+1} = G_k + hb(G_k) + \sqrt{h}\tilde{Z}_{k+1}$$

where $(\tilde{Z}_{k+1})_{k=0}^\infty$ are i.i.d. normally distributed random variables obtained through application of the transformation ψ . Note that if $G_0 \sim X_0$, then for all $k \geq 1$ we have $\mathcal{L}(G_k) = \mathcal{L}(X_k)$.

We now want to apply Theorem 2.5 with $Y' = G_{k+1}$, $y = G_k$, $X' = S_{k+1}$, $x = Y_{kh}$ and $\tilde{X} = Y_{(k+1)h}$. Hence for all $k \geq 0$ we obtain

$$\begin{aligned} \mathbb{E}f(|G_{k+1} - Y_{(k+1)h}|) &\leq (1 - ch)\mathbb{E}f(|G_k - Y_{kh}|) + \mathbb{E} \left[\frac{1}{r_2} e^{-ar_2} |S_{k+1} - Y_{(k+1)h}|^2 \right] \\ &\quad + \mathbb{E} \left[\left(1 + \frac{1}{r_2} e^{-ar_2} \left(|G_k - Y_{kh}|(1 + h_0 L) + \sqrt{h_0} \right) \right) |S_{k+1} - Y_{(k+1)h}| \right]. \end{aligned} \quad (2.32)$$

We need a few technical lemmas to bound the quantities appearing on the right hand side. All these lemmas work under Assumptions 1.1, with the exception of Lemma 2.13, where an additional upper bound on h is needed.

Lemma 2.11. *Assumptions 1.1 imply that for all $x \in \mathbb{R}^d$ we have*

$$\langle b(x), x \rangle \leq M_2 - M_1 |x|^2, \quad (2.33)$$

where

$$M_2 := L \left(\max \left(\mathcal{R}, \frac{2|b(0)|}{K} \right) \right)^2 + |b(0)| \max \left(\mathcal{R}, \frac{2|b(0)|}{K} \right) \text{ and } M_1 := \frac{K}{2}.$$

Lemma 2.12. *Let $(Y_t)_{t \geq 0}$ be defined by (2.24) and let b satisfy Assumptions 1.1. Then for any $t > 0$ we have*

$$\mathbb{E}|Y_t|^2 \leq C_{SDE},$$

where

$$C_{SDE} := \mathbb{E}|Y_0|^2 + \frac{2M_2 + d}{2M_1}.$$

Lemma 2.13. *Let $(X_k)_{k=0}^\infty$ be given by (2.25). Let $h < h_0 \wedge \frac{K}{4L^2}$. Then for all $k \geq 1$ we have*

$$\mathbb{E}|X_k|^2 \leq C_{Eul},$$

where

$$C_{Eul} := \mathbb{E}|X_0|^2 + \frac{2h_0|b(0)|^2 + d + M_2}{M_1 - 2h_0 L^2}. \quad (2.34)$$

Note that the bound obtained above applies also to the process $(G_k)_{k=0}^\infty$ defined by (2.31), since $\mathcal{L}(G_k) = \mathcal{L}(X_k)$ for all $k \geq 0$.

Lemma 2.14. *Let $(Y_{kh})_{k=0}^\infty$ and $(S_k)_{k=0}^\infty$ be defined by (2.26) and (2.30), respectively. Then for any $k \geq 0$ we have*

$$\mathbb{E}|S_{k+1} - Y_{(k+1)h}|^2 \leq C_{dif} h^3,$$

where

$$C_{dif} := L^2 \left(\frac{4h_0}{3} \left(\left(\mathbb{E}|Y_0|^2 + \frac{2M_2 + 1}{M_1} \right) L^2 + |b(0)|^2 \right) + d \right). \quad (2.35)$$

Proofs of all the lemmas can be found in the Appendix.

Combining all the estimates from the lemmas, we can come back to (2.32) and we see that

$$\begin{aligned} \mathbb{E}f(|G_{k+1} - Y_{(k+1)h}|) &\leq (1 - ch)\mathbb{E}f(|G_k - Y_{kh}|) + \frac{1}{r_2}e^{-ar_2}C_{dif}h^3 + \sqrt{C_{dif}}h^{3/2} \\ &\quad + \frac{1}{r_2}e^{-ar_2}(\sqrt{C_{Eul}} + \sqrt{C_{SDE}})(1 + h_0L)\sqrt{C_{dif}}h^{3/2} + \frac{1}{r_2}e^{-ar_2}\sqrt{h_0}\sqrt{C_{dif}}h^{3/2}. \end{aligned}$$

Let us define

$$C_{ult} := \frac{1}{r_2}e^{-ar_2}C_{dif}h_0^{3/2} + \sqrt{C_{dif}} + \frac{1}{r_2}e^{-ar_2}(\sqrt{C_{Eul}} + \sqrt{C_{SDE}})(1 + h_0L)\sqrt{C_{dif}} + \frac{1}{r_2}e^{-ar_2}\sqrt{h_0}\sqrt{C_{dif}}. \quad (2.36)$$

Then we have

$$\mathbb{E}f(|G_{k+1} - Y_{(k+1)h}|) \leq (1 - ch)\mathbb{E}f(|G_k - Y_{kh}|) + C_{ult}h^{3/2}$$

and hence

$$\begin{aligned} \mathbb{E}f(|G_{k+1} - Y_{(k+1)h}|) &\leq (1 - ch)^{k+1}\mathbb{E}f(|G_0 - Y_0|) + \sum_{j=0}^k (1 - ch)^j C_{ult}h^{3/2} \\ &\leq (1 - ch)^{k+1}\mathbb{E}f(|G_0 - Y_0|) + \frac{C_{ult}}{c}h^{1/2}. \end{aligned} \quad (2.37)$$

Using Lemma 2.3 we get

$$\left(\mathbb{E}|G_{k+1} - Y_{(k+1)h}|^2\right)^{1/2} \leq \left(A(1 - ch)^{k+1}\mathbb{E}f(|G_0 - Y_0|)\right)^{1/2} + \left(\frac{AC_{ult}}{c}\right)^{1/2}h^{1/4}.$$

Note that we could also use it to obtain L^1 bounds (cf. Remark 2.10)

$$\mathbb{E}|G_{k+1} - Y_{(k+1)h}| \leq e^{ar_2}(1 - ch)^{k+1}\mathbb{E}f(|G_0 - Y_0|) + e^{ar_2}\frac{C_{ult}}{c}h^{1/2}.$$

However, instead we will apply Theorems 2.6 and 2.7. Note that we have

$$\mathbb{E}f_1(|G_{k+1} - Y_{(k+1)h}|) \leq (1 - c_1h)\mathbb{E}f_1(|G_k - Y_{kh}|) + \mathbb{E}|S_{k+1} - Y_{(k+1)h}|$$

and $\mathbb{E}|S_{k+1} - Y_{(k+1)h}| \leq \sqrt{C_{dif}}h^{3/2}$ and hence

$$\mathbb{E}f_1(|G_{k+1} - Y_{(k+1)h}|) \leq (1 - c_1h)^{k+1}\mathbb{E}f_1(|G_0 - Y_0|) + \sum_{j=0}^k (1 - c_1h)^j \sqrt{C_{dif}}h^{3/2}.$$

Using $f_1(x) \geq e^{-qr_1^1}x$ for $x \in \mathbb{R}_+$ we get

$$\mathbb{E}|G_{k+1} - Y_{(k+1)h}| \leq e^{qr_1^1}(1 - c_1h)^{k+1}\mathbb{E}f_1(|G_0 - Y_0|) + e^{qr_1^1}\frac{\sqrt{C_{dif}}}{c_1}h^{1/2},$$

which finishes the proof. \square

Proof of Theorem 1.2. Since our reasoning in the proof of Theorem 2.9 applies for any coupling of the initial conditions X_0 and Y_0 , we can take the infimum on the right hand sides of (2.28) and (2.29) and thus we obtain upper bounds with $\mathbb{E}f(|X_0 - Y_0|)$ replaced by $W_f(\mathcal{L}(X_0), \mathcal{L}(Y_0))$. Hence both (1.8) and (1.9) follow immediately from Theorem 2.9 with $C_2 = \sqrt{A}$, $\hat{c}_2 = c$, $\tilde{c}_2 = \left(\frac{AC_{ult}}{c}\right)^{1/2}$, $C_1 = e^{qr_1^1}$, $\hat{c}_1 = c_1$ and $C_1 = e^{qr_1^1}\frac{\sqrt{C_{dif}}}{c_1}$. \square

2.4 Euler scheme with randomised (inaccurate) drift

Here we consider

$$\bar{X}_{k+1} = \bar{X}_k + h\bar{b}(\bar{X}_k, U_k) + \sqrt{h}Z_{k+1}$$

for $k \geq 0$ and compare it with the standard Euler scheme

$$X_{k+1} = X_k + hb(X_k) + \sqrt{h}Z_{k+1}. \quad (2.38)$$

We also note that since U_k and \bar{X}_k are independent, (1.11) implies

$$\mathbb{V}[\bar{b}(\bar{X}_k, U_k)|\bar{X}_k] \leq \sigma^2(1 + |\bar{X}_k|^2)h^\alpha \quad (2.39)$$

for any $k \geq 0$. Moreover, note that $\mathbb{E}[|\bar{b}(\bar{X}_k, U_k)|^2|\bar{X}_k] = \mathbb{V}[\bar{b}(\bar{X}_k, U_k)|\bar{X}_k] + |\bar{b}(\bar{X}_k)|^2$ and hence (1.4) and (2.39) imply

$$\mathbb{E}[|\bar{b}(\bar{X}_k, U_k)|^2|\bar{X}_k] \leq \sigma^2(|\bar{X}_k|^2 + 1)h^\alpha + 2L^2|\bar{X}_k|^2 + 2|b(0)|^2 \quad (2.40)$$

for any $k \geq 0$. These estimates will be used frequently in the sequel.

Now let us analyse an example that appears often in the statistics literature, see e.g. [WT11, Sha16] and the references therein, and explain how Assumption 1.3 can be verified.

Example 2.15. Let $(\theta_i)_{i=1,\dots,m}$ and $\theta_i \in \mathbb{R}^d$, for all i . Moreover, let $U = (U_i)_{i=1,\dots,s}$ be a collection of s independent random variables, uniformly distributed over the set $\{1, \dots, m\}$. We define

$$b(x) = \sum_{i=1}^m \hat{b}(x, \theta_i) \quad \text{and} \quad \bar{b}(x, U_k) = \frac{m}{s} \sum_{i=1}^s \hat{b}(x, \theta_{U_i}). \quad (2.41)$$

In applications of Bayesian inference m corresponds to the size of the data set and may be large. Consequently, the generation of $(\bar{X}_k)_{k=0}^\infty$ is costly. One then hopes that a randomisation strategy will reduce the computational cost without introducing significant variance. Notice that

$$\begin{aligned} \mathbb{E}[\bar{b}(x, U)] &= \frac{m}{s} \sum_{i=1}^s \mathbb{E}[\hat{b}(x, \theta_{U_i})] = \frac{m}{s} \sum_{i=1}^s \sum_{j=1}^m \hat{b}(x, \theta_j) \mathbb{P}(U_i = j) \\ &= \frac{1}{s} \sum_{i=1}^s \sum_{j=1}^m \hat{b}(x, \theta_j) = \sum_{i=1}^m \hat{b}(x, \theta_i) = b(x). \end{aligned}$$

On the other hand, we have

$$\begin{aligned} \mathbb{E}|\bar{b}(x, U) - b(x)|^2 &= \mathbb{E} \left| \frac{m}{s} \sum_{i=1}^s \hat{b}(x, \theta_{U_i}) - \sum_{i=1}^m \hat{b}(x, \theta_i) \right|^2 = \mathbb{E} \left| \frac{1}{s} \sum_{i=1}^s \left(m\hat{b}(x, \theta_{U_i}) - \sum_{i=1}^m \hat{b}(x, \theta_i) \right) \right|^2 \\ &= \frac{1}{s^2} \sum_{i=1}^s \mathbb{E} \left| m\hat{b}(x, \theta_{U_i}) - \sum_{i=1}^m \hat{b}(x, \theta_i) \right|^2 = \frac{1}{s^2} \sum_{i=1}^s \sum_{j=1}^m \left(m\hat{b}(x, \theta_j) - b(x) \right)^2 \frac{1}{m} \\ &= \frac{1}{s^2} \left(sm \sum_{j=1}^m \hat{b}(x, \theta_j)^2 - sb(x)^2 \right), \end{aligned}$$

where we used the fact that $m\hat{b}(x, \theta_{U_i}) - \sum_{i=1}^m \hat{b}(x, \theta_i)$ are centered, independent random variables. This implies that if for all θ and x we have $|\hat{b}(x, \theta)|^2 \leq C(1 + |x|^2)$ with some constant $C > 0$, then condition (1.11) can indeed be satisfied. Then we have

$$\mathbb{E}|\bar{b}(x, U) - b(x)|^2 \leq \frac{m^2}{s} C(1 + |x|^2)$$

and hence in order for (1.11) to be satisfied, we need to have $m^2 C/s \leq \sigma^2 h^\alpha$ for some constants $\sigma, \alpha > 0$, which means that we need to choose s and h so that $s^{-1} \lesssim h^\alpha$. However, this can give us a constant σ^2 of order m^2 , which can be very large in applications. In order to reduce the variance, a sensible choice seems to be to consider subsampling without replacement, see e.g. [Sha16]. More precisely, we define an estimator

$$\bar{b}^{wor}(x, U) := \frac{m}{s} \sum_{i=j}^m \hat{b}(x, \theta_j) Z_j,$$

where $(Z_j)_{j=1}^m$ are correlated random variables such that $\mathbb{P}(Z_j = 1) = \frac{s}{m}$, $\mathbb{P}(Z_j = 0) = 1 - \frac{s}{m}$ and $\mathbb{P}(Z_i = 1, Z_j = 1) = \frac{\binom{m-2}{s-2}}{\binom{m}{s}}$ for any $i, j \in \{1, \dots, m\}$ such that $i \neq j$. Note that this definition of \bar{b}^{wor} corresponds to sampling s terms from the sum defining b without replacement, see e.g. Lemma B in Section 7.3.1 and Problem 7.26 in [Ric88]. It is immediate to check that this estimator of b is indeed unbiased. As for the variance, we have

$$\mathbb{V} \left(\frac{m}{s} \sum_{i=j}^m \hat{b}(x, \theta_j) Z_j \right) = \frac{m^2}{s^2} \left(\sum_{j=1}^m \hat{b}(x, \theta_j)^2 \mathbb{V}(Z_j) + \sum_{j=1}^m \sum_{i \neq j} \hat{b}(x, \theta_j) \hat{b}(x, \theta_i) \text{Cov}(Z_i, Z_j) \right).$$

Moreover, it is easy to check that $\text{Cov}(Z_i, Z_j) = \frac{s(s-1)}{m(m-1)} - \frac{s^2}{m^2} = -\frac{s(1-\frac{s}{m})}{m(m-1)}$ and hence

$$\begin{aligned} \mathbb{V}(\bar{b}^{wor}(x, U)) &= \frac{m^2}{s^2} \left(\sum_{j=1}^m \hat{b}(x, \theta_j)^2 \frac{s}{m} (1 - \frac{s}{m}) - \sum_{j=1}^m \sum_{i \neq j} \hat{b}(x, \theta_j) \hat{b}(x, \theta_i) \frac{s(1 - \frac{s}{m})}{m(m-1)} \right) \\ &= \frac{m}{s} (1 - \frac{s}{m}) \frac{m}{m-1} \sum_{j=1}^m \left(\hat{b}(x, \theta_j) - \frac{b(x)}{m} \right)^2, \end{aligned}$$

where the last equality comes from the easily verifiable identity

$$\sum_{j=1}^m \left(\hat{b}(x, \theta_j) - \frac{b(x)}{m} \right)^2 = \frac{m-1}{m} \left(\sum_{j=1}^m \hat{b}(x, \theta_j)^2 - \frac{1}{m-1} \sum_{j=1}^m \sum_{i \neq j} \hat{b}(x, \theta_j) \hat{b}(x, \theta_i) \right).$$

Hence we see that the variance of the estimator \bar{b}^{wor} is equal to the variance of \bar{b} multiplied by $(1 - \frac{s}{m})$. Thus, assuming again that for all θ and x we have $|\hat{b}(x, \theta)|^2 \leq C(1 + |x|^2)$, we see that we now need to have

$$\frac{m^2}{s} (1 - \frac{s}{m}) C \leq \sigma^2 h^\alpha$$

in order for (1.11) to hold. Since the left hand side goes to zero when s approaches m , this method allows us to choose a much smaller σ than in the subsampling with replacement case, if we choose s large enough.

Another possible way of reducing the variance σ is via an appropriate rescaling (time-change) of the SDE (1.7). Namely, it is well-known that for any positive definite symmetric matrix Σ , the SDE $dY_t = \Sigma b(Y_t) dt + \Sigma^{1/2} dW_t$ has the same invariant measure as (1.7). See e.g. [DRVZ17, XSL⁺14] and the references therein for discussions on different choices of Σ in Monte Carlo algorithms. Hence, instead of considering the drift b and its estimator \bar{b} given by (2.41), we can take $b(x) = \frac{1}{m} \sum_{i=1}^m \hat{b}(x, \theta_i)$ and $\bar{b}(x, U_k) = \frac{1}{s} \sum_{i=1}^s \hat{b}(x, \theta_{U_i})$ and we can consider a Markov chain

$$\bar{X}_{k+1} = \bar{X}_k + h \frac{1}{s} \sum_{i=1}^s \hat{b}(\bar{X}_k, \theta_{U_i}) + \sqrt{h/m} Z_{k+1}.$$

This corresponds to choosing $\Sigma = (1/m)I$ in the SDE. Intuitively, for $m > 1$ this choice reduces the variance of the algorithm at the cost of slowing down convergence of the Markov chain. In other words, in our bounds in Theorem 1.4 the term $\bar{C}_2 h^{\alpha/4}$ becomes smaller, whereas $W_2(\mathcal{L}(X_k), \pi)$ becomes larger (since the contractivity constant \hat{c}_2 in Theorem 1.2 becomes smaller). A precise analysis of this trade-off falls beyond the scope of the present paper and is left for future work.

By applying Theorem 2.5, we can prove the following result.

Theorem 2.16. *Let the assumptions of Theorem 2.1 and Assumption 1.3 hold. Let $h \in (0, h_0 \wedge \frac{K}{4L^2+2\sigma^2} \wedge 1)$. Then for any random variables X_0, \bar{X}_0 and any $k \geq 1$ we have*

$$W_2(\mathcal{L}(X_k), \mathcal{L}(\bar{X}_k)) \leq \left(A(1 - ch)^k \mathbb{E}f(|X_0 - \bar{X}_0|) \right)^{1/2} + \left(\frac{AC_{I_{ult}}}{c} \right)^{1/2} h^{\alpha/4} \quad (2.42)$$

and

$$W_1(\mathcal{L}(X_k), \mathcal{L}(\bar{X}_k)) \leq e^{ar_2} (1 - ch)^k \mathbb{E}f(|X_0 - \bar{X}_0|) + e^{ar_2} \frac{C_{I_{ult}}}{c} h^{\alpha/2}, \quad (2.43)$$

where c is given by (2.13), A is given by (2.15), $C_{I_{ult}}$ is defined in (2.47) and α is specified in condition (1.11).

Proof. We will need to use an auxiliary chain $(S_k)_{k=0}^\infty$ that at each step k is moved to the position of \bar{X}_k and then evolves as (2.38), i.e.,

$$S_{k+1} := \bar{X}_k + hb(\bar{X}_k) + \sqrt{h}Z_{k+1}.$$

We define a new process $(G_k)_{k=0}^\infty$ coupled to $(S_k)_{k=0}^\infty$ according to the prescription (2.7). Using the notation introduced in (2.8), we set

$$G_{k+1} := \psi_{m,H}(\bar{X}_k, G_k, Z_{k+1}). \quad (2.44)$$

In other words, we have $G_{k+1} = G_k + hb(G_k) + \sqrt{h}\tilde{Z}_{k+1}$ for some i.i.d. normal random variables $(\tilde{Z}_k)_{k=1}^\infty$ that are determined via the coupling. Note that we have $\mathcal{L}(G_k) = \mathcal{L}(X_k)$ for all $k \geq 1$ if $G_0 = X_0$.

We apply now Theorem 2.5 and we have

$$\begin{aligned} \mathbb{E}f(|G_{k+1} - \bar{X}_{k+1}|) &\leq (1 - ch)\mathbb{E}f(|G_k - \bar{X}_k|) + \mathbb{E} \left[\frac{1}{r_2} e^{-ar_2} |S_{k+1} - \bar{X}_{k+1}|^2 \right] \\ &\quad + \mathbb{E} \left[\left(1 + \frac{1}{r_2} e^{-ar_2} \left(|G_k - \bar{X}_k|(1 + h_0 L) + \sqrt{h_0} \right) \right) |S_{k+1} - \bar{X}_{k+1}| \right]. \end{aligned} \quad (2.45)$$

Note that by (1.11) we have $\mathbb{E}|\bar{b}(\bar{X}_k, U_k) - b(\bar{X}_k)|^2 \leq \sigma^2(1 + \mathbb{E}|\bar{X}_k|^2)h^\alpha$ and hence, using Jensen's inequality,

$$\mathbb{E}|S_{k+1} - \bar{X}_{k+1}| = h\mathbb{E}|b(\bar{X}_k) - b(\bar{X}_k, U_k)| \leq \sigma h^{1+\alpha/2}(1 + \mathbb{E}|\bar{X}_k|^2)^{1/2}.$$

Moreover, we have uniform bounds on the moments of $(G_k)_{k=0}^\infty$ due to Lemma 2.13. Hence we only need to control the moments of $(\bar{X}_k)_{k=0}^\infty$. To this end, we can repeat the reasoning from the proof of Lemma 2.13 to obtain the following result.

Lemma 2.17. Let $(\bar{X}_k)_{k=0}^\infty$ be given by (1.10). Let $h < h_0 \wedge \frac{K}{4L^2+2\sigma^2} \wedge 1$. Then for all $k \geq 1$, under Assumptions 1.1 and 1.3, we have

$$\mathbb{E}|\bar{X}_k|^2 \leq C_{IEul},$$

where

$$C_{IEul} := \mathbb{E}|X_0|^2 + \frac{2h_0|b(0)|^2 + d + M_2 + h_0\sigma^2}{M_1 - 2h_0L^2 - h_0\sigma^2}. \quad (2.46)$$

The proof can be found in the Appendix. Now we come back to (2.45) and we have

$$\begin{aligned} \mathbb{E}f(|G_{k+1} - \bar{X}_{k+1}|) &\leq (1 - ch)\mathbb{E}f(|G_k - \bar{X}_k|) + \frac{1}{r_2}e^{-ar_2}\sigma^2h^{1+\alpha}(1 + C_{IEul}) \\ &\quad + \mathbb{E}\left(1 + \frac{1}{r_2}e^{-ar_2}\left((\sqrt{C_{Eul}} + \sqrt{C_{IEul}})(1 + h_0L) + \sqrt{h_0}\right)\right)\sigma h^{1+\alpha/2}(1 + C_{IEul})^{1/2}. \end{aligned}$$

Hence if we define

$$\begin{aligned} C_{Iult} &:= \frac{1}{r_2}e^{-ar_2}\sigma^2h_0^{\alpha/2}(1 + C_{IEul}) \\ &\quad + \left(\sigma + \sigma\frac{1}{r_2}e^{-ar_2}(\sqrt{C_{Eul}} + \sqrt{C_{IEul}})(1 + h_0L) + \sigma\frac{1}{r_2}e^{-ar_2}\sqrt{h_0}\right)(1 + C_{IEul})^{1/2}, \end{aligned} \quad (2.47)$$

we get

$$\mathbb{E}f(|G_{k+1} - \bar{X}_{k+1}|) \leq (1 - ch)\mathbb{E}f(|G_k - \bar{X}_k|) + C_{Iult}h^{1+\alpha/2}$$

and we can finish the proof as in the previous section, obtaining

$$\mathbb{E}f(|G_{k+1} - \bar{X}_{k+1}|) \leq (1 - ch)^{k+1}\mathbb{E}f(|G_0 - \bar{X}_0|) + \frac{C_{Iult}}{c}h^{\alpha/2}. \quad (2.48)$$

From this (2.42) and (2.43) follow easily due to Lemma 2.3. \square

Proof of Theorem 1.4. Since $W_2(\mathcal{L}(\bar{X}_k), \pi) \leq W_2(\mathcal{L}(\bar{X}_k), \mathcal{L}(X_k)) + W_2(\mathcal{L}(X_k), \pi)$ and an analogous inequality holds for the W_1 distance, it is easy to see that Theorem 2.16 with $\bar{X}_0 = X_0$ implies (1.12) and (1.13) with

$$\bar{C}_2 = \left(\frac{AC_{Iult}}{c}\right)^{1/2} \text{ and } \bar{C}_1 = e^{ar_2}\frac{C_{Iult}}{c},$$

respectively. \square

Remark 2.18. Note that the bound from Lemma 2.17 can be generalised to hold for all p -th moments of \bar{X}_k for $p \geq 1$. More precisely, for any $p \geq 1$ we can prove that under conditions (1.14) and (1.15), there exists a constant $C_{IEul}^{(p)} > 0$ such that for sufficiently small h and for all $k \geq 1$ we have

$$\mathbb{E}|\bar{X}_k|^p \leq C_{IEul}^{(p)}. \quad (2.49)$$

In order to see this, we first need to show an analogous bound for the Euler scheme (2.25) with accurate drift as in Lemma 2.13, under (1.14). This follows e.g. from Theorem 2.1 and Remark 2.4 in [SZ18]. Indeed, note that (1.14) implies that for any $p \geq 2$ we have $\langle x, b(x) \rangle |x|^{p-2} \leq M_2|x|^{p-2} - M_1|x|^p$ and hence for the generator L of the solution to the SDE (1.7) for the function $V(x) = |x|^p$ we can show

$$\begin{aligned} LV(x) &= \langle \nabla V(x), b(x) \rangle + \frac{1}{2}\Delta V(x) = p|x|^{p-2}\langle x, b(x) \rangle + \frac{1}{2}p(p-1)|x|^{p-2} \\ &\leq pM_2|x|^{p-2} - pM_1|x|^p + \frac{1}{2}p(p-1)|x|^{p-2} \\ &\leq -\left(pM_1 - \varepsilon pM_2 - \frac{1}{2}\varepsilon p(p-1)\right)|x|^p + \left(pM_2 + \frac{1}{2}p(p-1)\right)C(\varepsilon, p), \end{aligned}$$

where $\varepsilon > 0$ can be arbitrary and $C(\varepsilon, p) > 0$ is such that $|x|^{p-2} \leq \varepsilon|x|^p + C(\varepsilon, p)$ for all $x \in \mathbb{R}^d$ (i.e., we can choose $C(\varepsilon, p) > \left(\frac{p-2}{\varepsilon p}\right)^{\frac{p-2}{2}} - \varepsilon \left(\frac{p-2}{\varepsilon p}\right)^{\frac{p}{2}}$). Hence, if we choose ε such that $\varepsilon < M_1/(M_2 + \frac{1}{2}(p-1))$, we obtain a Lyapunov condition $LV(x) \leq -\rho V(x) + C$ with some positive constants $\rho, C > 0$. Observe that Theorem 2.1 and Remark 2.4 in [SZ18] imply that under a stronger condition, namely, $LV(x) \leq -\rho(1 + V(x))$ for all $x \in \mathbb{R}^d$, (cf. (2.8) in [SZ18]), we get (2.49) for Euler schemes with accurate drifts. However, by analysing the proof in [SZ18] it is easy to see that their condition (2.8) can be replaced by the weaker condition $LV(x) \leq -\rho V(x) + C$. Namely, in the last calculation in the proof of Theorem 2.1 in [SZ18], the inequality $\mathbb{E}(1 + V(\bar{X}_{k+1})) \leq (1 + (-\rho + \tilde{\rho})h)\mathbb{E}(1 + V(\bar{X}_k))$ will be then replaced by $\mathbb{E}(1 + V(\bar{X}_{k+1})) \leq (1 + (-\rho + \tilde{\rho})h)\mathbb{E}(1 + V(\bar{X}_k)) + (C + \rho)h$. Then by iterating we obtain $\mathbb{E}(1 + V(\bar{X}_{k+1})) \leq e^{(-\rho + \tilde{\rho})(k+1)h}\mathbb{E}(1 + V(X_0)) + C + \rho$ and the desired bound for the accurate drift case follows. Now, using (1.15) and following the argument from the proof of Lemma 2.17, it is possible to extend this result to Euler schemes with inaccurate drifts. We leave the details to the reader.

2.5 Multi-level Monte Carlo

In this section we focus on Euler schemes with inaccurate drifts. However, the reader who is interested only in MLMC in the accurate drift case, can easily recover relevant results by replacing $(\bar{X}_k)_{k=0}^\infty$ defined in (1.10) with $(X_k)_{k=0}^\infty$ defined in (1.1) and Assumptions 1.6 with Assumptions 1.1. Note that in such a case certain quantities featured below simply vanish, which only makes our calculations easier.

Let us start by briefly explaining the motivation behind considering Monte Carlo estimators of the multi-level type. A typical strategy for approximating $\int_{\mathbb{R}^d} g(x)\pi(dx)$, is to resort to the standard Monte Carlo estimator where the average is taken "over the space". More precisely, we fix the time $T = kh$ for some $k \geq 1$, we generate N i.i.d. samples $(\bar{X}_k^i)_{i=1}^N$ of \bar{X}_k defined in (1.10) and compute the Monte Carlo average

$$\mathcal{A}^{MCA}(T, h, N)(g) := \frac{1}{N} \sum_{i=1}^N g(\bar{X}_k^i). \quad (\text{MCA})$$

The aim is to find the optimal allocation of the parameters (terminal time T , number of MC samples N and the size of the time-step h) to achieve required mean-square-error. We can compute

$$\begin{aligned} mse(\mathcal{A}^{MCA}(T, h, N)(g)) &:= \left[\mathbb{E} \left(\int_{\mathbb{R}^d} g(x)\pi(dx) - \frac{1}{N} \sum_{i=1}^N g(\bar{X}_k^i) \right)^2 \right]^{1/2} \\ &\leq \left[\left(\int_{\mathbb{R}^d} g(x)(\pi(dx) - \mu_{kh}(dx)) \right)^2 \right]^{1/2} + \left[(\mathbb{E}[g(Y_{kh})] - \mathbb{E}[g(\bar{X}_k)])^2 \right]^{1/2} \\ &\quad + \left[\mathbb{E} \left(\mathbb{E}[g(\bar{X}_k)] - \frac{1}{N} \sum_{i=1}^N g(\bar{X}_k^i) \right)^2 \right]^{1/2}, \end{aligned}$$

where $\mu_{kh} := \mathcal{L}(Y_{kh})$. The three error terms are: bias (due to finite time simulation) that we can estimate due to [Ebe16] with explicit constants if g is Lipschitz, i.e., $W_1(\mathcal{L}(Y_{kh}), \pi) \leq e^{-\lambda kh} W_1(\mathcal{L}(Y_0), \pi)$ or due to [LW16] if g has polynomial growth; weak time discretisation error studied in Theorem 1.5 from which we know that $|\mathbb{E}[g(Y_{kh})] - \mathbb{E}[g(\bar{X}_k)]| \lesssim h$, and the variance of the Monte Carlo estimator that we also control uniformly in time due to Lemma 2.17. Hence

we have $\mathcal{A}^{MCA}(T, h, N) \lesssim e^{-\lambda T} + h + 1/\sqrt{N}$. We fix $\epsilon > 0$ and set $\mathcal{A}^{MCA}(T, h, N) \lesssim \epsilon$. This leads to the following choice of the parameters $T \approx \lambda^{-1} \log \epsilon$, $h \approx \epsilon$, $N \approx \epsilon^{-2}$. The computational cost is then given by $\text{cost}(\mathcal{A}^{MCA}(t, h, N)) = Th^{-1}N \approx (\log \epsilon)\epsilon^{-3}$. The above cost should be compared with ϵ^{-2} that holds for the MC estimator in the case when we have access to unbiased samples.

The recently developed MLMC approach, [Gil08, Keb05, Hei01], allows us to reduce the computational cost of (MCA). The idea is to introduce a family of Euler discretisations with varying time-steps. Fix $L > 0$. For $\ell \in \{1, \dots, L\}$ let us define $h^\ell := \frac{1}{M^\ell}$. In our analysis we consider $M = 2$, for simplicity. We define

$$\bar{X}_{(k+1)h^\ell}^\ell = \bar{X}_{kh^\ell}^\ell + \bar{b}(\bar{X}_{kh^\ell}^\ell, U_{kh^\ell}^\ell)h^\ell + Z_{k+1}^\ell, \quad (2.50)$$

where $Z_{k+1}^\ell := W_{(k+1)h^\ell} - W_{kh^\ell}$ and $(U_{kh^\ell}^\ell)_{k=0}^\infty$ are mutually independent and such that for any $x \in \mathbb{R}^d$ we have $\mathbb{E}\bar{b}(x, U_{kh^\ell}^\ell) = b(x)$. For all $\ell \in \{1, \dots, L\}$ and all T such that $T = kh^\ell$ for some $k \geq 1$, we introduce appropriate modifications of (2.50) denoted by $(\bar{X}_T^{f,\ell}, \bar{X}_T^{c,\ell})$ such that $\mathcal{L}(\bar{X}_T^{f,\ell}) = \mathcal{L}(\bar{X}_T^{c,\ell}) = \mathcal{L}(\bar{X}_T)$ for $\ell \in \{1, \dots, L\}$. Hence we have

$$\mathbb{E}[g(\bar{X}_T^L)] = \mathbb{E}[g(\bar{X}_T^0)] + \sum_{\ell=1}^L \mathbb{E}[g(\bar{X}_T^{f,\ell}) - g(\bar{X}_T^{c,\ell-1})]. \quad (2.51)$$

This identity leads to the following unbiased estimator of $\mathbb{E}[g(\bar{X}_T^L)]$, for $T = kh^L$ with some $k \geq 1$,

$$\mathcal{A}^{MLMC}(T, L, (N_\ell)_\ell)(g) := \frac{1}{N_0} \sum_{i=1}^{N_0} g(\bar{X}_T^{i,0}) + \sum_{\ell=1}^L \left\{ \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} (g(\bar{X}_T^{i,f,\ell}) - g(\bar{X}_T^{i,c,\ell-1})) \right\},$$

where $\bar{X}_T^{i,f,\ell}$ and $\bar{X}_T^{i,c,\ell}$ for $i \in \{1, \dots, N_\ell\}$ are i.i.d. samples of $\bar{X}_T^{f,\ell}$ and $\bar{X}_T^{c,\ell}$, respectively. We assume that the samples across the levels (each summand in ℓ) are independent, hence

$$\text{Var}(\mathcal{A}^{MLMC}(T, L, (N_\ell)_\ell)(g)) = \frac{1}{N_0} \text{Var}(g(\bar{X}_T^{1,0})) + \sum_{\ell=1}^L \frac{1}{N_\ell} \text{Var}(g(\bar{X}_T^{i,f,\ell}) - g(\bar{X}_T^{i,c,\ell-1})). \quad (2.52)$$

If the samples at each summand in ℓ , i.e., $(\bar{X}_T^{i,f,\ell}, \bar{X}_T^{i,c,\ell-1})$ are appropriately coupled for each ℓ , then $\text{Var}(g(\bar{X}_T^{i,f,\ell}) - g(\bar{X}_T^{i,c,\ell-1}))$ decays when ℓ increases. As a result MLMC, with optimally selected parameters, combines many simulations on low accuracy grids (at a corresponding low cost), with relatively few simulations computed with high accuracy and high cost on very fine grids. One has the following error decomposition

$$\begin{aligned} \text{mse}(\mathcal{A}^{MLMC}(T, L, (N_\ell)_\ell)(g)) &= \left[\mathbb{E} \left(\int_{\mathbb{R}^d} g(x) \pi(dx) - \mathcal{A}^{MLMC}(T, L, (N_\ell)_\ell)(g) \right)^2 \right]^{1/2} \\ &\leq \left[\left(\int_{\mathbb{R}^d} g(x) (\pi(dx) - \mu_T(dx)) \right)^2 \right]^{1/2} + \left[\left(\mathbb{E}[g(Y_T)] - \mathbb{E}[g(\bar{X}_T^{1,L})] \right)^2 \right]^{1/2} \\ &\quad + \left[\left(\mathbb{E}[g(\bar{X}_T^{1,L})] - \mathcal{A}^{MLMC}(T, L, (N_\ell)_\ell)(g) \right)^2 \right]^{1/2}. \end{aligned}$$

Note that the first two errors are the same as for the standard MC and the only difference will come from the variance of the MLMC estimator. In order to evoke to the classical Multilevel

Monte Carlo complexity analysis, note that we applied telescopic sum estimator to $\mathbb{E}[g(Y_T)]$ rather than directly to $\int_{\mathbb{R}^d} g(x)\pi(dx)$. In view of the analysis for the standard Monte Carlo we choose $T \approx \lambda^{-1} \log \epsilon$ and hence we need to multiply the final cost of MLMC by the factor $\log \epsilon$.

Observe that in analysis of an MLMC estimator the crucial part is to investigate the behaviour of the pair of processes $(\bar{X}_{kh}^{f,\ell}, \bar{X}_{kh}^{c,\ell-1})_{k=0}^\infty$ for any fixed $\ell \in \{1, \dots, L\}$. Hence, to streamline the notation, from now on we drop the superscript ℓ and we will work with

$$\bar{X}_{(k+1)h}^f = \bar{X}_{kh}^f + \bar{b}(\bar{X}_{kh}^f, U_{kh}^f)h + \sqrt{h}Z_{k+1}, \quad \bar{X}_{(k+2)h}^f = \bar{X}_{(k+1)h}^f + \bar{b}(\bar{X}_{(k+1)h}^f, U_{(k+1)h}^f)h + \sqrt{h}Z_{k+2} \quad (2.53)$$

and

$$\bar{X}_{(k+2)h}^c = \bar{X}_{kh}^c + \bar{b}(\bar{X}_{kh}^c, U_{kh}^c)2h + \sqrt{2h}\hat{Z}_{k+2}, \quad (2.54)$$

where $\hat{Z}_{k+2} := (Z_{k+1} + Z_{k+2})/\sqrt{2}$. Thus $(\bar{X}_{kh}^f)_{k=0}^\infty$ is a process on a fine grid and $(\bar{X}_{kh}^c)_{k=0}^\infty$ is a process on a coarse grid that moves twice less frequently than $(\bar{X}_{kh}^f)_{k=0}^\infty$.

For the condition with the telescopic sum (2.51) to hold, it is required that for all $k \in \mathbb{N}$, $\mathcal{L}(U_{kh}^f) = \mathcal{L}(U_{kh}^c)$. Taking as an example subsampling considered in Example 2.15, we see that this condition forces us to take the same number of samples at each step of the algorithm. We also assume that $(U_{kh}^f)_{k=0}^\infty$ are mutually independent so that (2.52) holds. The random variables $(U_{kh}^c)_{k=0}^\infty$ can be chosen as independent of $(U_{kh}^f)_{k=0}^\infty$, although coupling them in an appropriate way can help to further reduce the variance (see Remark 2.21).

We impose the following assumptions.

Assumptions 2.19. We assume that

- Random variables $(U_{kh}^f)_{k \in \mathbb{N}}$ and $(U_{kh}^c)_{k \in \mathbb{N}}$ are such that for all $k \in \mathbb{N}$, $\mathcal{L}(U_{kh}^f) = \mathcal{L}(U_{kh}^c)$.
- There are constants $L_u > 0$ and $\alpha_c > 0$ such that for all $x \in \mathbb{R}^d$ we have

$$\mathbb{E}|\bar{b}(x, U_{kh}^f) + \bar{b}(x, U_{(k+1)h}^f) - 2\bar{b}(x, U_{kh}^c)|^2 \leq L_u(1 + |x|^2)h^{\alpha_c}, \quad \forall k \geq 0. \quad (2.55)$$

We have the following result.

Theorem 2.20. Suppose Assumptions 1.3, 1.6 with $\bar{L} = L$, $\bar{K} = K$, $\bar{R} = \mathcal{R}$, and 2.19 are satisfied. Then there exists a process $(\bar{G}_k)_{k=0}^\infty$ such that $\mathcal{L}(\bar{G}_k) = \mathcal{L}(\bar{X}_{kh}^f)$ for all $k \geq 1$ if $\bar{G}_0 = \bar{X}_0^f$ and we have

$$\mathbb{E}|\bar{G}_k - \bar{X}_{kh}^c|^2 \leq A(1 - ch)^k \mathbb{E}f(|\bar{G}_0 - \bar{X}_0^c|) + \frac{A \cdot C_{IMLult}}{c} h^{\min(\alpha_c, 1)/2}$$

for all $h \in (0, h_0 \wedge \frac{K}{4L^2 + 2\sigma^2} \wedge 1)$, where the constant C_{IMLult} is defined in (2.58), c is given by (2.13) and A is given by (2.15).

This result allows us to construct an MLMC estimator with L levels and good variance bounds by using couplings $(\bar{G}_k, \bar{X}_{kh}^c)_{k=0}^\infty$. Namely, we start by considering the coarsest level $\ell = 0$, we take $\bar{X}_{kh}^c = \bar{X}_k^0$ and define the process $(\bar{G}_k)_{k=0}^\infty$ as explained in the proof of Theorem 2.20, which corresponds to the finer level $\ell = 1$. Then we treat thus obtained $(\bar{G}_k)_{k=0}^\infty$ as a new coarse process and we repeat our procedure $L - 1$ times.

Remark 2.21. Assumptions 2.19 can be easily satisfied by choosing the random variables U_{kh}^f , $U_{(k+1)h}^f$ and U_{kh}^c independently and then using condition (1.11) to verify (2.55). Then we obtain $\alpha_c = \alpha$. However, we would like to point out that such an approach to getting a bound on the variance of our MLMC estimator is not necessarily optimal. We believe that using more involved couplings of $U_{(k+1)h}^f$, U_{kh}^f and U_{kh}^c could lead to $\alpha_c > \alpha$ and hence to an improvement of the rate that we obtain above. We will consider this problem in our future research.

Proof of Theorem 1.7. It is shown in Giles [Gil15] that under the assumptions

$$|\mathbb{E}[g(Y_T) - g(\bar{X}_T^{1,\ell})]| \lesssim (h^\ell)^{\tilde{\alpha}}, \quad \text{Var}[g(\bar{X}_T^{1,\ell}) - g(\bar{X}_T^{1,\ell-1})] \lesssim (h^\ell)^\beta, \quad (2.56)$$

for some $\tilde{\alpha} \geq 1/2$, $\beta > 0$, the computational complexity of the resulting multi-level estimator with accuracy ε is proportional to

$$\mathcal{C} = \begin{cases} \varepsilon^{-2}, & \beta > \gamma, \\ \varepsilon^{-2} \log^2(\varepsilon), & \beta = \gamma, \\ \varepsilon^{-2-(1-\beta)/\tilde{\alpha}}, & 0 < \beta < \gamma \end{cases}$$

where the cost of sampling at level ℓ is of order $h^{-\ell\gamma}$. Theorem 1.5 tells us that in our case $\tilde{\alpha} = 1$ while Theorem 2.20 with $\bar{G}_0 = \bar{X}_0^c$ gives $\beta = \min(\alpha_c, 1)/2$ since g is Lipschitz. The cost of sampling at level ℓ is proportional to $(h^\ell)^{-1}$, hence $\gamma = 1$. Hence an overall cost of approximating $\mathbb{E}[g(Y_T)]$ is of order $\varepsilon^{-2-(1-\min(\alpha_c, 1)/2)}$. Consequently, the cost of approximating $\int_{\mathbb{R}^d} g(x)\pi(dx)$ is $\varepsilon^{-2-(1-\min(\alpha_c, 1)/2)} \log \varepsilon$. In particular, if we verify condition (2.55) by using (1.11) as explained in Remark 2.21, we have $\alpha_c = \alpha$ and the cost is $\varepsilon^{-2-(1-\min(\alpha, 1)/2)} \log \varepsilon$. Hence the best gain that we can get is when we manage to choose $\alpha_c = 1$ (or $\alpha = 1$), which gives us $\beta = 1/2$ and the cost $\varepsilon^{-5/2} \log \varepsilon$, which is a half order of magnitude better than the standard Monte Carlo approach. \square

Remark 2.22. In [GMS⁺19], authors considered a telescopic sum in two parameters, time-discretisation and time t (length of the chain). We could also apply this idea in the current setting, which would lead to the reduction of overall cost by $\log \varepsilon$.

Remark 2.23. As our primary interest in the present paper was to study the general randomised (inaccurate) Euler scheme, we refrain from comparing the complexity of the subsampling algorithms described in Example 2.15 with their multi-level counterparts. In that example one would need to take into consideration additional cost of simulating a step of Euler scheme (m for accurate gradient and s for inaccurate gradient). We leave the details to the interested reader.

2.5.1 Variance bound for the inaccurate drift case

Proof of Theorem 2.20. We consider the processes $(\bar{X}_{kh}^f)_{k=0}^\infty$ and $(\bar{X}_{2kh}^c)_{k=0}^\infty$ defined via (2.53) and (2.54), respectively. We will obtain a bound on the variance by applying Theorem 2.8 to $\bar{X}_{(k+2)h}^f$ and $\bar{X}_{(k+2)h}^c$. We consider the process $(\bar{S}_{kh})_{k=0}^\infty$ defined for $k \in 2\mathbb{N}$ by

$$\begin{aligned} \bar{S}_{(k+1)h} &= \bar{X}_{kh}^c + h\bar{b}(\bar{X}_{kh}^c, U_{kh}^f) + \sqrt{h}Z_{k+1} \\ \bar{S}_{(k+2)h} &= \bar{S}_{(k+1)h} + h\bar{b}(\bar{S}_{(k+1)h}, U_{(k+1)h}^f) + \sqrt{h}Z_{k+2}. \end{aligned}$$

We also need the process $(\bar{G}_k)_{k=0}^\infty$ coupled to $(\bar{S}_{kh})_{k=0}^\infty$ via

$$\bar{G}_{k+1} := \bar{\psi}_{m,H}(\bar{G}_k, \bar{S}_{kh}, U_{kh}^f, Z_{k+1}). \quad (2.57)$$

for all $k \in \mathbb{N}$, where we use the notation from (2.23). Then we have $\mathcal{L}(\bar{G}_k) = \mathcal{L}(\bar{X}_{kh}^f)$ for all $k \geq 1$ if $\bar{G}_0 = \bar{X}_0^f$. Applying Theorem 2.8, we obtain

$$\begin{aligned} \mathbb{E}f(|\bar{G}_{k+2} - \bar{X}_{(k+2)h}^c|) &\leq (1 - ch)\mathbb{E}f(|\bar{G}_{k+1} - \bar{S}_{(k+1)h}|) + \mathbb{E} \left[\frac{1}{r_2} e^{-ar_2} |\bar{S}_{(k+2)h} - \bar{X}_{(k+2)h}^c|^2 \right] \\ &\quad + \mathbb{E} \left[\left(1 + \frac{1}{r_2} e^{-ar_2} \left(|\bar{G}_{k+1} - \bar{S}_{(k+1)h}|(1 + h_0 L) + \sqrt{h_0} \right) \right) |\bar{S}_{(k+2)h} - \bar{X}_{(k+2)h}^c| \right]. \end{aligned}$$

Note now that

$$\begin{aligned} & \mathbb{E} \left[\left(1 + \frac{1}{r_2} e^{-ar_2} \left(|\bar{G}_{k+1} - \bar{S}_{(k+1)h}|(1 + h_0 L) + \sqrt{h_0} \right) \right) |\bar{S}_{(k+2)h} - \bar{X}_{(k+2)h}^c| \right] \\ &= \left(1 + \frac{1}{r_2} e^{-ar_2} \sqrt{h_0} \right) \mathbb{E} |\bar{S}_{(k+2)h} - \bar{X}_{(k+2)h}^c| \\ &+ \frac{1}{r_2} e^{-ar_2} (1 + h_0 L) \mathbb{E} \left[|\bar{G}_{k+1} - \bar{S}_{(k+1)h}| \cdot |\bar{S}_{(k+2)h} - \bar{X}_{(k+2)h}^c| \right] =: I_1 + I_2 \end{aligned}$$

and we have

$$I_2 \leq \frac{1}{r_2} e^{-ar_2} (1 + h_0 L) \left(\mathbb{E} |\bar{G}_{k+1} - \bar{S}_{(k+1)h}|^2 \right)^{1/2} \cdot \left(\mathbb{E} |\bar{S}_{(k+2)h} - \bar{X}_{(k+2)h}^c|^2 \right)^{1/2}.$$

Now we use Minkowski's inequality and Lemma 2.17 to estimate

$$\left(\mathbb{E} |\bar{G}_{k+1} - \bar{S}_{(k+1)h}|^2 \right)^{1/2} \leq \left(\mathbb{E} |\bar{G}_{k+1}|^2 \right)^{1/2} + \left(\mathbb{E} |\bar{S}_{(k+1)h}|^2 \right)^{1/2} \leq \sqrt{C_{IEul}} + \left(\mathbb{E} |\bar{S}_{(k+1)h}|^2 \right)^{1/2}.$$

Moreover, due to (1.11) we have

$$\mathbb{E} |\bar{b}(\bar{X}_{kh}^c, U_{kh}^f)|^2 \leq \sigma^2 (\mathbb{E} |\bar{X}_{kh}^c|^2 + 1) h^\alpha + 2L^2 \mathbb{E} |\bar{X}_{kh}^c|^2 + 2|b(0)|^2 \leq C_{IEul}^{(2h)} (\sigma^2 + 2L^2) + \sigma^2 + 2|b(0)|^2$$

where we used $h^\alpha \leq 1$ and we have the constant

$$C_{IEul}^{(2h)} := \mathbb{E} |\bar{X}_0^c|^2 + \frac{4h_0 |b(0)|^2 + d + M_2 + 2h_0 \sigma^2}{M_1 - 4h_0 L^2 - 2h_0 \sigma^2}$$

which is specified as in (2.46) but with h replaced by $2h$. If we denote

$$C_{ISM} := C_{IEul}^{(2h)} (\sigma^2 + 2L^2) + \sigma^2 + 2|b(0)|^2$$

then we get

$$\begin{aligned} \left(\mathbb{E} |\bar{S}_{(k+1)h}|^2 \right)^{1/2} &= \left(\mathbb{E} |\bar{X}_{kh}^c + h\bar{b}(\bar{X}_{kh}^c, U_{kh}^f) + \sqrt{h}Z_{k+1}|^2 \right)^{1/2} \\ &\leq \left(\mathbb{E} |\bar{X}_{kh}^c|^2 \right)^{1/2} + \left(h^2 \mathbb{E} |\bar{b}(\bar{X}_{kh}^c, U_{kh}^f)|^2 \right)^{1/2} + \left(h \mathbb{E} |Z_{k+1}|^2 \right)^{1/2} \\ &\leq \sqrt{C_{IEul}^{(2h)}} + h_0 \sqrt{C_{ISM}} + \sqrt{h_0 d}. \end{aligned}$$

Hence, denoting $C_{IASP} := \left(\sqrt{C_{IEul}^{(2h)}} + h_0 \sqrt{C_{ISM}} + \sqrt{h_0 d} \right)^2$ we obtain

$$I_2 \leq \frac{1}{r_2} e^{-ar_2} (1 + h_0 L) (\sqrt{C_{IEul}} + \sqrt{C_{IASP}}) \left(\mathbb{E} |\bar{S}_{(k+2)h} - \bar{X}_{(k+2)h}^c|^2 \right)^{1/2}.$$

Now we estimate

$$\begin{aligned} \mathbb{E} |\bar{S}_{(k+2)h} - \bar{X}_{(k+2)h}^c|^2 &= \mathbb{E} |\bar{X}_{kh}^c + h\bar{b}(\bar{X}_{kh}^c, U_{kh}^f) + \sqrt{h}Z_{k+1} \\ &+ h\bar{b}(\bar{X}_{kh}^c + h\bar{b}(\bar{X}_{kh}^c, U_{kh}^f) + \sqrt{h}Z_{k+1}, U_{(k+1)h}^f) + \sqrt{h}Z_{k+2} - \bar{X}_{kh}^c - 2h\bar{b}(\bar{X}_{kh}^c, U_{kh}^c) - \sqrt{2h}\hat{Z}_{k+2}|^2 \\ &\leq 2\mathbb{E} |h\bar{b}(\bar{X}_{kh}^c, U_{kh}^f) + h\bar{b}(\bar{X}_{kh}^c, U_{(k+1)h}^f) - 2h\bar{b}(\bar{X}_{kh}^c, U_{kh}^c)|^2 \\ &+ 2\mathbb{E} |h\bar{b}(\bar{X}_{kh}^c + h\bar{b}(\bar{X}_{kh}^c, U_{kh}^f) + \sqrt{h}Z_{k+1}, U_{(k+1)h}^f) - h\bar{b}(\bar{X}_{kh}^c, U_{(k+1)h}^f)|^2 =: \tilde{I}_1 + \tilde{I}_2. \end{aligned}$$

Note that due to (1.16) we have

$$\tilde{I}_2 \leq 2h^2 \bar{L}^2 \mathbb{E} |h\bar{b}(\bar{X}_{kh}^c, U_{kh}^f) + \sqrt{h}Z_{k+1}|^2 \leq 4h^4 \bar{L}^2 C_{ISM} + 4h^3 \bar{L}^2 d.$$

and due to (2.55) we have

$$\tilde{I}_1 \leq 2h^2 L_u (1 + \mathbb{E} |\bar{X}_{kh}^c|^2) h^{\alpha_c} \leq 2L_u (1 + C_{IEul}^{(2h)}) h^{2+\alpha_c}.$$

If we denote

$$C_{IMLdif} := (4h_0 \bar{L}^2 C_{ISM} + 4\bar{L}d) h_0^{(1-\alpha_c)^+} + 2L_u (1 + C_{IEul}^{(2h)}),$$

then we see that $\mathbb{E} |\bar{S}_{(k+2)h} - \bar{X}_{(k+2)h}^c|^2 \leq C_{IMLdif} h^{2+\min(\alpha_c, 1)}$ and we have

$$I_2 \leq \frac{1}{r_2} e^{-ar_2} (1 + h_0 L) (\sqrt{C_{IEul}} + \sqrt{C_{IASP}}) \sqrt{C_{IMLdif}} h^{1+\min(\alpha_c, 1)/2}$$

and

$$I_1 \leq \left(1 + \frac{1}{r_2} e^{-ar_2} \sqrt{h_0}\right) \sqrt{C_{IMLdif}} h^{1+\min(\alpha_c, 1)/2},$$

This gives us

$$\mathbb{E} f(|\bar{G}_{k+2} - \bar{X}_{(k+2)h}^c|) \leq (1 - ch) \mathbb{E} f(|\bar{G}_{k+1} - \bar{S}_{(k+1)h}|) + C_{IMLult} h^{1+\min(\alpha_c, 1)/2},$$

where

$$\begin{aligned} C_{IMLult} &:= \frac{1}{r_2} e^{-ar_2} C_{IMLdif} h_0^{1+\min(\alpha_c, 1)/2} + \left(1 + \frac{1}{r_2} e^{-ar_2} \sqrt{h_0}\right) \sqrt{C_{IMLdif}} \\ &\quad + \frac{1}{r_2} e^{-ar_2} (1 + h_0 L) (\sqrt{C_{IEul}} + \sqrt{C_{IASP}}) \sqrt{C_{IMLdif}}. \end{aligned} \quad (2.58)$$

Now we just apply the contractivity result for processes with inaccurate drifts (Theorem 2.8) to \bar{G}_{k+1} and $\bar{S}_{(k+1)h}$ (note that they are appropriately coupled due to (2.57)) and we obtain

$$\mathbb{E} f(|\bar{G}_{k+2} - \bar{X}_{(k+2)h}^c|) \leq (1 - ch)^2 \mathbb{E} f(|\bar{G}_k - \bar{S}_{(k)h}|) + C_{IMLult} h^{1+\min(\alpha_c, 1)/2}.$$

Hence for all $k \in 2\mathbb{N}$ we obtain

$$\begin{aligned} \mathbb{E} f(|\bar{G}_k - \bar{X}_{kh}^c|) &\leq (1 - ch)^k \mathbb{E} f(|\bar{G}_0 - \bar{X}_0^c|) + \sum_{j=0}^{k-1} (1 - ch)^j C_{IMLult} h^{1+\min(\alpha_c, 1)/2} \\ &\leq (1 - ch)^k \mathbb{E} f(|\bar{G}_0 - \bar{X}_0^c|) + \frac{C_{IMLult}}{c} h^{\min(\alpha_c, 1)/2}. \end{aligned}$$

Using Lemma 2.3 we obtain

$$\mathbb{E} |\bar{G}_k - \bar{X}_{kh}^c|^2 \leq A(1 - ch)^k \mathbb{E} f(|\bar{G}_0 - \bar{X}_0^c|) + \frac{AC_{IMLult}}{c} h^{\min(\alpha_c, 1)/2}$$

with the constant A given by (2.15), which finishes the proof. \square

3 Proof of the contractivity result

As we have already remarked in Section 2.2, we only need to prove Theorem 2.8 and hence we will be working here with inaccurate drifts with randomness induced by a random variable U independent of Z . Recall that we have $\hat{x} = x + h\bar{b}(x, U)$ and $\hat{y} = y + h\bar{b}(y, U)$.

Note that some parts of this proof (Lemmas 3.1, 3.2 and Subsection 3.3) are based on Section 6 in [EM19]. We include all the details here for the reader's convenience and in order to highlight the modifications that need to be introduced in the proofs from [EM19] in order to obtain L^2 bounds. The reader who is only interested in Euler schemes with non-random drifts can obtain a direct proof of Theorem 2.1 from the reasoning below by setting $\hat{x} = x + hb(x)$ and replacing $\mathbb{E}[\cdot | U]$ with $\mathbb{E}[\cdot]$ and the assumptions (1.16) and (1.17) with (1.4) and (1.5), respectively.

We denote $R' := |X' - Y'|$, $\hat{r} := |\hat{x} - \hat{y}|$ and $r := |x - y|$. We want to show that for any $r \in [0, \infty)$ we have

$$\mathbb{E}[f(R') | U] - f(r) \leq -chf(r) \text{ a.s.} \quad (3.1)$$

Before we proceed with the detailed proof, let us sketch some main ideas and formulate two crucial lemmas.

We proceed by decomposing the Euler scheme step $r \mapsto R'$ into the drift step $r \mapsto \hat{r}$ and the Gaussian step $\hat{r} \mapsto R'$. We have $\mathbb{E}[f(R') | U] - f(r) = \mathbb{E}[f(R') | U] - f(\hat{r}) + f(\hat{r}) - f(r)$ and we will want to use either $\mathbb{E}[f(R') | U] - f(\hat{r})$ or $f(\hat{r}) - f(r)$ (depending on the values of r and \hat{r}) to get an upper bound of the form $-chf(r)$.

The main idea is that whenever we are in the region of space where the contractivity condition (1.17) holds, we should use the drift step and switch off the Gaussian movement by applying the synchronous coupling. On the other hand, when the contractivity condition does not work, we have to use an appropriate coupling to get the desired upper bounds from the Gaussian step, while controlling the drift step via the Lipschitz condition (1.16).

One of the most important parts of the proof is an application of the Taylor formula to obtain

$$f(R') - f(\hat{r}) = f'(\hat{r})(R' - \hat{r}) + f''(\theta)(R' - \hat{r})^2$$

for some θ between \hat{r} and R' . Hence we see that our crucial task is to control the first and the second moment of $R' - \hat{r}$. Actually, we will be able to consider the second moment only when R' is restricted to a specific interval. In fact we will choose random (only through dependence on U) intervals

$$I_{\hat{r}} = \begin{cases} (0, \hat{r} + \sqrt{h}) & \text{if } \hat{r} \leq \sqrt{h} \\ (\hat{r} - \sqrt{h}, \hat{r}) & \text{if } \hat{r} > \sqrt{h}. \end{cases} \quad (3.2)$$

For such intervals, we will obtain the random variable $\underline{\alpha}(\hat{r})$, for which we have

$$\underline{\alpha}(\hat{r}) \leq \mathbb{E}[(R' - \hat{r})^2 \mathbf{1}_{\{R' \in I_{\hat{r}}\}} | U] \text{ a.s.}$$

The reason for this choice of intervals $I_{\hat{r}}$ will become apparent from the proof of Lemma 3.2. Generally speaking, we want to choose a small interval around \hat{r} and it is convenient to take $(\hat{r} - \sqrt{h}, \hat{r})$ for getting bounds on $\sup f''$ on $[0, r_1]$. However, we cannot take just $(0, \hat{r})$ if $\hat{r} \leq \sqrt{h}$ as the length of such an interval would not have a uniform lower bound and it would be impossible to get the lower bound in Lemma 3.2, hence we need to take $(0, \hat{r} + \sqrt{h})$ when \hat{r} is small.

From now on, all the relations between random variables in the proof are supposed to be understood as holding almost surely. Let us first formulate two auxiliary results on our coupling, which allow us to control the conditional moments of $R' - \hat{r}$.

Lemma 3.1. For the coupling (X', Y') defined by (2.23), we have $\mathbb{E}[R' | U] = \hat{r}$.

Lemma 3.2. For the coupling (X', Y') defined by (2.23), if $h \leq 4m^2$, then we have

$$\underline{\alpha}(\hat{r}) \mathbf{1}_{\{\hat{r} \leq H\}} \leq \mathbb{E}[(R' - \hat{r})^2 \mathbf{1}_{\{R' \in I_{\hat{r}}\}} | U],$$

where $\underline{\alpha}(\hat{r}) = \frac{1}{2}c_0 \min(\sqrt{h}, \hat{r})\sqrt{h}$ and

$$c_0 := 4 \min \left(\int_0^{1/2} u^2 (1 - e^{u-1/2}) \varphi_{0,1}(u) du, (1 - e^{-1}) \int_0^{1/2} u^3 \varphi_{0,1}(u) du \right).$$

The proofs are based on the calculations from Section 6 in [EM19] and can be found in Appendix 8.

Before we proceed, let us make a remark about the choice of parameters in our coupling (2.7).

Remark 3.3. The choice of $m = \sqrt{h_0}/2$ in (2.7) means that we rarely make the non-synchronous step. In principle, we could choose m arbitrarily large, but then from the proof we see that we would also need to redefine $r_2 := r_1 + 2m$, and as r_2 increases, the constant c decreases to zero. Moreover, we see that $A \rightarrow \infty$ as $r_2 \rightarrow \infty$. This shows that increasing m actually gives us worse constants and hence m should be kept as small as possible. This is related to the fact that the only place in the calculations where we gain something from the non-synchronous behaviour of our coupling is in the lower bound for $\underline{\alpha}$, which is the coefficient near f'' . But this supremum is taken over a small interval and actually when we compute the lower bound for $\underline{\alpha}$ we only integrate a Gaussian density over an interval of length $[0, \sqrt{h}/2]$. Everywhere else the non-synchronous behaviour is actually harmful to our estimates (due to the convex cost), so it makes sense to take m very small, i.e., $m = \frac{\sqrt{h_0}}{2}$ and $r_2 = r_1 + 2m = r_1 + \sqrt{h_0}$.

We are now ready to proceed with the proof of Theorem 2.8, which we will split into a few steps, depending on the size of the argument r . First note that if $r = |x - y| > \mathcal{R}$, then due to (1.16) and (1.17) we have

$$\begin{aligned} \hat{r} &= \sqrt{r^2 + 2h\langle x - y, \bar{b}(x, U) - \bar{b}(y, U) \rangle + h^2|\bar{b}(x, U) - \bar{b}(y, U)|^2} \\ &\leq \sqrt{1 - 2hK + h^2L^2} \\ &\leq r \left(1 - hK + \frac{h^2L^2}{2} \right), \end{aligned}$$

where we used the fact that $\sqrt{1+x} \leq 1 + \frac{x}{2}$ for all $x \geq -1$. Note that due to Assumptions 1.1 we have $K \leq L$ and hence the expression under the square root is indeed non-negative. Therefore if we choose h small enough, we have $\hat{r} \leq r$ for $r > \mathcal{R}$. More precisely, we will assume $h \leq K/L^2$, which implies

$$\frac{h^2L^2}{2} = \frac{hL^2}{2}h \leq \frac{hL^2}{2} \frac{K}{L^2} = \frac{hK}{2},$$

hence in fact we have $-hK + \frac{h^2L^2}{2} \leq -\frac{hK}{2}$ and we will later use the fact that for all $r > \mathcal{R}$ we have

$$\hat{r} \leq r \left(1 - \frac{hK}{2} \right). \quad (3.3)$$

However, even if $r \leq \mathcal{R}$, we can still control the size of \hat{r} due to (1.16). Namely, we have

$$\hat{r} = \sqrt{|x - y + h(\bar{b}(x, U) - \bar{b}(y, U))|^2} \leq (1 + hL)\mathcal{R} \leq (1 + h_0L)\mathcal{R} = r_1$$

for $r \leq \mathcal{R}$. This is the condition motivating the choice of r_1 , i.e., if r is in $[0, \mathcal{R}]$, then the application of the drift can increase the distance \hat{r} maximally up to r_1 .

Having the above basic estimates, we can begin the proof of the easiest case, when $r \in [r_1, r_2]$.

3.1 Estimates for $r \in [r_1, r_2]$.

Since $r \geq r_1 > \mathcal{R}$, we have $\hat{r} \leq r \leq r_2$. This means that the interval (\hat{r}, r) is contained in $[0, r_2]$, where the function f is concave. Hence for the drift step we have

$$f(\hat{r}) - f(r) \leq f'(r)(\hat{r} - r). \quad (3.4)$$

Furthermore, the derivative of f is bounded from below on the interval $[0, r_2]$ by $f'(r_2) = e^{-ar_2}$, which implies

$$-\frac{Khr}{2}f'(r) \leq -\frac{Khr}{2}e^{-ar_2} \leq -\frac{Kh}{2}e^{-ar_2}f(r), \quad (3.5)$$

where in the last step we additionally use the fact that $f(r) \leq r$ for $r \in [0, r_2]$. Combining (3.3), (3.4) and (3.5) gives $f(\hat{r}) - f(r) \leq -\frac{Kh}{2}e^{-ar_2}f(r)$ for all $r \in [r_1, r_2]$, which is exactly what we want. Now we have to show that the Gaussian step does not spoil these estimates. We need to consider two cases.

The case of $\hat{r} \leq r_1$. Observe that due to our coupling construction, the value of R' is always within the interval $[0, \hat{r} + 2m]$ (it can become zero when we jump to the same point, and if we reflect the jumps then R' can increase maximally up to $\hat{r} + 2m$ due to the truncation of jumps by m). Hence the interval with endpoints \hat{r} and R' is always contained in $[0, \hat{r} + 2m]$. Hence, if $\hat{r} \leq r_1$, then $(\hat{r}, R') \subset [0, r_1 + 2m] = [0, r_2]$, where (\hat{r}, R') should be interpreted as (R', \hat{r}) if $R' < \hat{r}$. Recalling that the function f is concave on $[0, r_2]$, by the Taylor formula we get

$$\mathbb{E}[f(R') | U] - f(\hat{r}) = f'(\hat{r})\mathbb{E}[(R' - \hat{r}) | U] + \mathbb{E}[f''(\theta)(R' - \hat{r})^2 | U] \leq f'(\hat{r})\mathbb{E}[(R' - \hat{r}) | U]$$

for some $\theta \in (\hat{r}, R') \subset [0, r_2]$. However, due to Lemma 3.1 we have $\mathbb{E}[(R' - \hat{r}) | U] = 0$ and hence $\mathbb{E}[f(R') | U] - f(\hat{r}) \leq 0$.

The case of $\hat{r} > r_1$. Here the interval $[0, \hat{r} + 2m]$ is no longer contained in $[0, r_2]$ and the function f is convex for arguments larger than r_2 , so we cannot bound f'' by zero as above. This is why for $\hat{r} > r_1$ we use the synchronous coupling and we have $\mathbb{E}[f(R') | U] - f(\hat{r}) = 0$.

Hence, combining all the estimates from this subsection together, we see that we have $\mathbb{E}[f(R') | U] - f(r) \leq -\frac{Kh}{2}e^{-ar_2}f(r)$ for all $r \in [r_1, r_2]$.

3.2 Estimates for $r \in [r_2, \infty)$.

Here we deal with the Gaussian step exactly as in the previous subsection. If $\hat{r} \leq r_1$, then we can use concavity of f on $[0, r_2]$. Otherwise, for $\hat{r} > r_1$, we use the synchronous coupling. Hence $\mathbb{E}[f(R') | U] - f(\hat{r}) \leq 0$ and we only need to deal with the drift step. Since $r \geq r_2 > \mathcal{R}$, we can bound \hat{r} by using (3.3). However, we do not know whether $r(1 - \frac{Kh}{2})$ is smaller or greater than r_2 , and since the function f is given via two different formulas for arguments below and above r_2 , we need to consider two different cases.

The case of $r(1 - \frac{Kh}{2}) \geq r_2$. Since f is increasing, using (3.3) we have

$$f(\hat{r}) \leq f\left(r\left(1 - \frac{Kh}{2}\right)\right) = \frac{1}{a}(1 - e^{-ar_2}) + \frac{1}{2r_2}e^{-ar_2}\left(r^2\left(1 - \frac{Kh}{2}\right)^2 - r_2^2\right). \quad (3.6)$$

We want to bound this expression from above by $f(r) - c'hf(r)$ for some constant $c' > 0$ and we know that, since $r \geq r_2$, we have

$$f(r) = \frac{1}{a}(1 - e^{-ar_2}) + \frac{1}{2r_2}e^{-ar_2}(r^2 - r_2^2). \quad (3.7)$$

A simple calculation shows that we need to find $c' > 0$ such that

$$\frac{1}{2r_2}e^{-ar_2}\left(-Kh + \frac{K^2h^2}{4}\right)r^2 \leq -c'hf(r) \quad (3.8)$$

holds for all $r \geq r_2$. In order for this to be possible, the left hand side needs to be negative. Note that $-Kh + \frac{K^2h^2}{4} < 0$ when $h < \frac{4}{K}$, which holds due to our choice of h_0 , cf. (2.9). We have

$$\begin{aligned} \left(Kh - \frac{K^2h^2}{4}\right)f(r) &= \left(Kh - \frac{K^2h^2}{4}\right)\frac{1}{2r_2}e^{-ar_2}r^2 \\ &+ \left(Kh - \frac{K^2h^2}{4}\right)\left(\frac{1}{a}(1 - e^{-ar_2}) - \frac{1}{2r_2}e^{-ar_2}r_2^2\right). \end{aligned} \quad (3.9)$$

Now observe that we can find a constant $c_2 > 0$ such that

$$\frac{1}{a}(1 - e^{-ar_2}) - \frac{1}{2r_2}e^{-ar_2}r_2^2 \leq c_2\frac{1}{2r_2}e^{-ar_2}r^2 \quad (3.10)$$

for all $r \geq r_2$. Namely, we set

$$c_2 := \frac{\frac{1}{a}(1 - e^{-ar_2}) - \frac{1}{2r_2}e^{-ar_2}r_2^2}{\frac{1}{2r_2}e^{-ar_2}r_2^2}.$$

We know that $c_2 > 0$ since the function $g(r) := \frac{1}{a}(1 - e^{-ar}) - \frac{1}{2}e^{-ar}r$ is increasing for all $r > 0$ and $g(0) = 0$. From (3.9) and (3.10) we get

$$\frac{\left(Kh - \frac{K^2h^2}{4}\right)}{1 + c_2}f(r) \leq \left(Kh - \frac{K^2h^2}{4}\right)\frac{1}{2r_2}e^{-ar_2}r^2$$

for all $r \geq r_2$. Combining this with (3.8), we obtain

$$f(\hat{r}) \leq f(r) - \frac{\left(Kh - \frac{K^2h^2}{4}\right)}{1 + c_2}f(r) = f(r) - \frac{\frac{1}{2}e^{-ar_2}r_2}{\frac{1}{a}(1 - e^{-ar_2})}\left(Kh - \frac{K^2h^2}{4}\right)f(r).$$

The case of $r(1 - \frac{Kh}{2}) \leq r_2$. We again use the fact that $\hat{r} \leq r(1 - \frac{Kh}{2})$ and $f(\hat{r}) \leq f(r(1 - \frac{Kh}{2}))$. The difficulty in this case comes from the fact that we need to compare the values of f given by two different formulas, i.e., $f(r)$ given by (3.7) and $f(r(1 - \frac{Kh}{2})) = \frac{1}{a}(1 - e^{-a(r(1 - \frac{Kh}{2}))})$. We can circumvent this problem by considering the midpoint between $r(1 - \frac{Kh}{2})$ and r , namely, the point $r(1 - \frac{Kh}{4})$. We have

$$f(\hat{r}) \leq f\left(r\left(1 - \frac{Kh}{2}\right)\right) = f\left(r\left(1 - \frac{Kh}{2}\right)\right) - f\left(r\left(1 - \frac{Kh}{4}\right)\right) + f\left(r\left(1 - \frac{Kh}{4}\right)\right). \quad (3.11)$$

We have either $r(1 - \frac{Kh}{4}) \leq r_2$ or $r(1 - \frac{Kh}{4}) > r_2$. In the former case, we use the fact that f is increasing to get $f(r(1 - \frac{Kh}{4})) \leq f(r)$ and for the remaining term in (3.11) we proceed similarly to Subsection 3.1. Namely, we have

$$\begin{aligned} f\left(r\left(1 - \frac{Kh}{2}\right)\right) - f\left(r\left(1 - \frac{Kh}{4}\right)\right) &\leq f'\left(r\left(1 - \frac{Kh}{4}\right)\right) \left(r\left(1 - \frac{Kh}{2}\right) - r\left(1 - \frac{Kh}{4}\right)\right) \\ &= -\frac{Kh}{4} f'\left(r\left(1 - \frac{Kh}{4}\right)\right) r \end{aligned}$$

due to concavity of f on $[0, r_2]$ and the fact that $[r(1 - \frac{Kh}{2}), r(1 - \frac{Kh}{4})] \subset [0, r_2]$. Now we use the fact that $f'(r) \geq e^{-ar_2}$ and $f(r) \leq r$ for $r \in [0, r_2]$, which gives us

$$-\frac{Kh}{4} f'\left(r\left(1 - \frac{Kh}{4}\right)\right) r \leq -e^{-ar_2} r \frac{Kh}{4} \leq -e^{-ar_2} \frac{Kh}{4} f(r).$$

On the other hand, if $r(1 - \frac{Kh}{4}) > r_2$, then we use the fact that f is increasing to get $f(r(1 - \frac{Kh}{2})) - f(r(1 - \frac{Kh}{4})) \leq 0$ and we have $f(\hat{r}) - f(r) \leq f(r(1 - \frac{Kh}{4})) - f(r)$. Since $r(1 - \frac{Kh}{4}) > r_2$, we can use our calculations from Subsection 3.2. We just need to replace $(1 - \frac{Kh}{2})$ therein with $(1 - \frac{Kh}{4})$, and, in consequence, we obtain

$$f\left(r\left(1 - \frac{Kh}{4}\right)\right) - f(r) \leq -\frac{\frac{1}{2}e^{-ar_2}r_2}{\frac{1}{a}(1 - e^{-ar_2})} \left(\frac{Kh}{2} - \frac{K^2h^2}{16}\right) f(r).$$

Note that we have $\frac{K}{2} - \frac{K^2h}{16} > \frac{K}{4}$ since $h < \frac{4}{K}$ by (2.9). Hence, combining all the estimates from this subsection together, we see that for all $r \in [r_2, \infty)$ we have

$$\mathbb{E}[f(R') | U] - f(r) \leq f(\hat{r}) - f(r) \leq -\min\left(e^{-ar_2} \frac{Kh}{4}, \frac{\frac{1}{2}e^{-ar_2}r_2}{\frac{1}{a}(1 - e^{-ar_2})} \frac{Kh}{4}\right) f(r).$$

This is also valid for $r \in [r_1, r_2]$, cf. the constant obtained in Subsection 3.1.

3.3 Estimates for $r \in [0, r_1]$.

For the drift step we have two cases. If $\hat{r} \leq r$, then we just use the fact that f is increasing and we get $f(\hat{r}) - f(r) \leq 0$. If $\hat{r} > r$, we will need to use $f(\hat{r}) - f(r) \leq f'(\hat{r})(\hat{r} - r)$, which holds due to the fact that $r \leq r_1$ implies $\hat{r} \leq r_1$ and f is concave on $[0, r_1]$. We will show later in this section how to bound this term in order to be able to use it in connection with the Gaussian step contribution to obtain the desired bounds.

For the Gaussian step, we again use that $r \in [0, r_1]$ implies $\hat{r} \in [0, r_1]$. Therefore, by our coupling construction, $R' \in [0, r_1 + 2m] = [0, r_2]$ and we can use the fact that f'' is concave on $[0, r_2]$. We apply the Taylor formula (note that the function f is twice differentiable), i.e., we have

$$\mathbb{E}[f(R') | U] - f(\hat{r}) = f'(\hat{r})\mathbb{E}[(R' - \hat{r}) | U] + \mathbb{E}[f''(\theta)(R' - \hat{r})^2 | U],$$

where θ is a point between \hat{r} and R' . We bound

$$\begin{aligned}
\mathbb{E} [f''(\theta)(R' - \hat{r})^2 | U] &\leq \mathbb{E} \left[\sup_{\theta \in (\hat{r}, R')} f''(\theta)(R' - \hat{r})^2 \mid U \right] \\
&= \mathbb{E} \left[\sup_{\theta \in (\hat{r}, R')} f''(\theta)(R' - \hat{r})^2 \mathbf{1}_{\{R' \in I_{\hat{r}}\}} + \sup_{\theta \in (\hat{r}, R')} f''(\theta)(R' - \hat{r})^2 \mathbf{1}_{\{R' \notin I_{\hat{r}}\}} \mid U \right] \\
&\leq \mathbb{E} \left[\sup_{\theta \in I_{\hat{r}}} f''(\theta)(R' - \hat{r})^2 \mathbf{1}_{\{R' \in I_{\hat{r}}\}} + \sup_{\theta \in (\hat{r}, R')} f''(\theta)(R' - \hat{r})^2 \mathbf{1}_{\{R' \notin I_{\hat{r}}\}} \mid U \right] \\
&\leq \sup_{\theta \in I_{\hat{r}}} f''(\theta) \mathbb{E} [(R' - \hat{r})^2 \mathbf{1}_{\{R' \in I_{\hat{r}}\}} | U] ,
\end{aligned}$$

where (\hat{r}, R') denotes either (\hat{r}, R') or (R', \hat{r}) , depending on the relation between \hat{r} and R' , $I_{\hat{r}}$ can be any (random) open interval such that \hat{r} belongs to the closure of $I_{\hat{r}}$, in the third line we use the fact that $R' \in I_{\hat{r}}$ implies that $(\hat{r}, R') \subset I_{\hat{r}}$ and in the last step we use the fact that $f'' \leq 0$ on $[0, r_2]$ so we can bound the second term by zero, whereas in the first expression $\sup_{\theta \in I_{\hat{r}}} f''(\theta)$ is pulled out in front of the conditional expectation as it is a measurable function of U .

We also use Lemma 3.1 to get $\mathbb{E} [(R' - \hat{r}) | U] = 0$. Hence we have

$$\mathbb{E} [f(R') | U] - f(\hat{r}) \leq \frac{1}{2} \underline{\alpha}(\hat{r}) \sup_{I_{\hat{r}}} f'' , \quad (3.12)$$

where $\underline{\alpha}(\hat{r}) \leq \mathbb{E} [(R' - \hat{r})^2 \mathbf{1}_{\{R' \in I_{\hat{r}}\}} | U]$ and the exact value of $\underline{\alpha}(\hat{r})$ is specified in Lemma 3.2 (note that here $\hat{r} \leq r_1 = H$) for our particular choice of intervals $I_{\hat{r}}$ given by (3.2). Recall that for the drift step we get

$$f(\hat{r}) - f(r) \leq f'(r)(\hat{r} - r) \quad (3.13)$$

and that due to (1.16) we have $\hat{r} = |x - y + h(\bar{b}(x, U) - \bar{b}(y, U))| \leq |x - y| + h|\bar{b}(x, U) - \bar{b}(y, U)| \leq r + hLr$. We would like to replace $f'(r)$ in (3.13) with $f'(\hat{r})$, since in (3.12) we consider supremum of f'' on a small neighbourhood of \hat{r} rather than r . Recall now that $f'(r) = e^{-ar}$ for $r \leq r_2$ (here we consider $r \leq r_1 \leq r_2$) and hence $f'(r) = e^{-a(r - \hat{r} + \hat{r})} = e^{-a\hat{r}} e^{-a(r - \hat{r})} = f'(\hat{r}) e^{-a(r - \hat{r})}$. We also have $\hat{r} - r \leq hLr$ and hence we can bound

$$f(\hat{r}) - f(r) \leq f'(r)(\hat{r} - r) \leq f'(\hat{r}) e^{ahLr_1} (\hat{r} - r) \leq f'(\hat{r}) e^{ahLr_1} hL\hat{r} ,$$

where the last step holds for $r \leq \hat{r}$. Note, however, that for $r > \hat{r}$ we have $f(\hat{r}) - f(r) \leq 0$ and hence the bound

$$f(\hat{r}) - f(r) \leq f'(\hat{r}) e^{ahLr_1} hL\hat{r} \quad (3.14)$$

holds for all $r \in [0, r_1]$. Thus, combining (3.12) and (3.14), we see that we need to show that

$$f'(\hat{r}) e^{ahLr_1} hL\hat{r} + \frac{1}{2} \underline{\alpha}(\hat{r}) \sup_{I_{\hat{r}}} f'' \leq -c' h f(r)$$

for some $c' > 0$ and for all $r \in [0, r_1]$. We will in fact show

$$f'(\hat{r}) e^{ahLr_1} hL\hat{r} + \frac{1}{2} \underline{\alpha}(\hat{r}) \sup_{I_{\hat{r}}} f'' \leq -ch f(\hat{r}) . \quad (3.15)$$

Note that this implies an upper bound by $-ch f(r)/2$ as long as $hL \in (0, 1/2)$. Indeed, we have $r = |x - y| \leq \hat{r} + hLr$ and thus $f(\hat{r}) \geq f((1 - hL)r) \geq (1 - hL)f(r) \geq f(r)/2$, since f is increasing and concave on $[0, r_1]$ and $hL \in (0, 1/2)$ by assumption (2.9).

Recall now that

$$I_{\hat{r}} = \begin{cases} (0, \hat{r} + \sqrt{h}) & \text{if } \hat{r} \leq \sqrt{h} \\ (\hat{r} - \sqrt{h}, \hat{r}) & \text{if } \hat{r} > \sqrt{h} \end{cases}$$

and f'' is increasing on $[0, r_2]$, since $f'''(r) = a^2 e^{-ar}$ for $r \in [0, r_2]$. Therefore

$$\sup_{I_{\hat{r}}} f'' = \begin{cases} -ae^{-a(\hat{r} + \sqrt{h})} & \text{if } \hat{r} \leq \sqrt{h} \\ -ae^{-a\hat{r}} & \text{if } \hat{r} > \sqrt{h} \end{cases}.$$

We proceed by considering two separate cases, related to the form of the interval $I_{\hat{r}}$ and, consequently, the form of the bound $\underline{\alpha}(\hat{r})$ from Lemma 3.2.

The case of $\hat{r} > \sqrt{h}$. Looking at (3.15), we see that we need to show that

$$e^{-a\hat{r}} e^{ahLr_1} hL\hat{r} - \frac{1}{2} \underline{\alpha}(\hat{r}) a e^{-a\hat{r}} \leq -ch \frac{1}{a} (1 - e^{-a\hat{r}}) \quad (3.16)$$

for all $\hat{r} \leq r_1$. Recall that $\underline{\alpha}(\hat{r}) = c_0 \min(\hat{r}, \sqrt{h}) \sqrt{h}$. We will show that

$$e^{-a\hat{r}} e^{ahLr_1} L\hat{r} - \frac{1}{2} c_0 a e^{-a\hat{r}} \leq -c \frac{1}{a} (1 - e^{-a\hat{r}}). \quad (3.17)$$

To this end it is sufficient to show that the left hand side of (3.17) is bounded by $-c/a$ for all $\hat{r} \in [0, r_1]$. Hence we need

$$e^{ahLr_1} L\hat{r} - \frac{1}{2} c_0 a \leq -\frac{c}{a} e^{a\hat{r}}.$$

Recall that $r_1 = (1 + h_0 L) \mathcal{R} \leq 3\mathcal{R}/2$, since $h_0 L < 1/2$ by assumption (2.9). Since $h < h_0$ and we have

$$h_0 \leq \frac{2c_0 \ln \frac{3}{2}}{27L^2 \mathcal{R}^2} \leq \frac{c_0 \ln \frac{3}{2}}{9L^2 r_1 \mathcal{R}} \leq \frac{\ln \frac{3}{2}}{aLr_1},$$

we see that $e^{ahLr_1} \leq 3/2$. We would like to stress the fact that this bound (and hence the assumption on h_0) is to some extent arbitrary and we can modify it in order to have in the last estimate any number strictly greater than 1 instead of $3/2$. We need to have

$$\frac{3}{2} L\hat{r} - \frac{1}{2} c_0 a \leq -\frac{c}{a} e^{a\hat{r}}. \quad (3.18)$$

First we need to make sure that $3L\hat{r} - c_0 a < 0$ holds for all $\hat{r} \leq r_1$, by choosing a in an appropriate way. This is guaranteed by our choice of $a = 6Lr_1/c_0$ made in (2.10), i.e., it is chosen (again to some extent arbitrarily, up to a constant) so that $\frac{3}{2} Lr_1 = \frac{1}{4} c_0 a$. With this choice of a we have $\frac{3}{2} L\hat{r} - \frac{1}{2} c_0 a \leq -\frac{1}{4} c_0 a$ for all $\hat{r} \leq r_1$ and hence in order to get (3.18) we need to have $-\frac{1}{4} c_0 a \leq -\frac{c}{a} e^{a\hat{r}}$ for all $\hat{r} \leq r_1$, which implies that we should choose

$$c \leq \frac{9L^2 r_1^2}{c_0} e^{-\frac{6Lr_1^2}{c_0}}.$$

The case of $\hat{r} \leq \sqrt{h}$. We use the fact that $f(\hat{r}) \leq \hat{r}$ and in fact for small \hat{r} these quantities are close to each other. Coming back to (3.16), we see that we need to show

$$e^{ahLr_1} hL - \frac{1}{2} c_0 \sqrt{h} a e^{-a\sqrt{h}} \leq -che^{a\hat{r}}. \quad (3.19)$$

Since $h < h_0$ and by assumption

$$h_0 \leq \frac{c_0^2 (\ln 2)^2}{144L^2 \mathcal{R}^2},$$

using once again the bound $r_1 \leq 3\mathcal{R}/2$, we see that

$$\sqrt{h} \leq \frac{c_0 \ln 2}{12L\mathcal{R}} \leq \frac{3c_0 \ln 2}{24Lr_1} = \frac{3}{4a} \ln 2 \leq \frac{1}{a} \ln 2$$

and we have $e^{a\sqrt{h}} \leq 2$. Hence the left hand side of (3.19) is bounded (recall that we also have $e^{ahLr_1} \leq \frac{3}{2}$) by $\frac{3}{2}hL - \frac{1}{4}c_0\sqrt{h}a$, which we now need to make sure is negative. We can do this e.g. by making sure that $\frac{3}{2}hL \leq \frac{1}{8}c_0\sqrt{h}a$, which is equivalent to $\sqrt{h} \leq \frac{r_1}{2}$, but this is ensured by our assumption $h_0 \leq \mathcal{R}^2/4$, since $\mathcal{R} \leq r_1$. Hence we see that the left hand side of (3.19) is bounded from above by $-c_0a\sqrt{h}/8$. Observe that we have $e^{a\hat{r}} \leq e^{a\sqrt{h}} \leq 2$, which implies $-\frac{1}{8}c_0a\sqrt{h} \leq -\frac{1}{16}c_0a\sqrt{h}e^{a\hat{r}}$. Hence we need to have

$$-\frac{1}{16}c_0a\sqrt{h}e^{a\hat{r}} \leq -che^{a\hat{r}},$$

which implies that we should choose

$$c \leq \frac{3Lr_1}{8\sqrt{h_0}} = \frac{c_0a}{16\sqrt{h_0}}.$$

Then (3.19) is satisfied and the proof is finished.

4 Weak error analysis

Proof of Theorem 1.5. The novelty of this proof is in the fact that we are working with an inaccurate drift. This is what we focus on, while keeping some standard estimates left to the reader. We refer to [Tal90] and [MT04] for details of proofs of weak convergence rates. Moreover, in order to clearly showcase the main ideas, we present the proof only in the one-dimensional setting. An extension to the multidimensional case follows in the same way, although with more cumbersome notation. Let $u(t, x) := \mathbb{E}[g(Y_t^{0,x})]$, where $Y_t^{0,x}$ is a solution to the SDE (1.7) at time t , started from the initial condition $x \in \mathbb{R}$ at time 0. Since $b \in C_b^3$ and $g \in C^\infty$ with polynomial growth, we can deduce from [Kry99] that $u \in C^{1,3}([0, T], \mathbb{R})$ and that it satisfies the Kolmogorov equation

$$\begin{cases} (Lu)(t, x) := \partial_t u(t, x) - \partial_x u(t, x)b(x) - \frac{1}{2}\partial_x^2 u(t, x) = 0, & (t, x) \in [0, \infty) \times \mathbb{R}, \\ u(0, x) = g(x). \end{cases} \quad (4.1)$$

Furthermore, due to our assumption (1.14), Theorem 3.4 in [Tal90] tells us that for any multi-index I with $|I| \leq 2$ there exist an integer $p_{|I|}$ and positive constants $\gamma_{|I|}$ and $\Gamma_{|I|}$ such that

$$|\partial_I u(t, x)| \leq \Gamma_{|I|}(1 + |x|^{p_{|I|}})e^{-\gamma_{|I|}t}. \quad (4.2)$$

Let $T = Mh$ for some $M \geq 1$. We define the continuous extension of (1.10), given by

$$\bar{X}_t = X_0 + \int_0^t \bar{b}(\bar{X}_{\eta(s)}, U_{\eta(s)})ds + W_t,$$

where $\eta(s) = kh$ for $s \in (kh, (k+1)h]$ and $(U_{kh})_{k=0}^\infty$ are defined in an analogous way as $(U_k)_{k=0}^\infty$ in (1.10). It is important to note that \bar{X}_{kh} is independent of U_{kh} . From the definition of u we

have $u(0, \bar{X}_T) = g(\bar{X}_T)$. Consequently $\mathbb{E}[g(\bar{X}_T)] - \mathbb{E}[g(Y_T)] = \mathbb{E}[u(0, \bar{X}_T) - u(T, X_0)]$ due to the Feynman-Kac formula. Then the Itô formula gives

$$\begin{aligned} \mathbb{E}[g(\bar{X}_T)] - \mathbb{E}[g(Y_T)] &= \sum_{k=1}^M \mathbb{E}[u(T - kh, \bar{X}_{kh}) - u(T - (k-1)h, \bar{X}_{(k-1)h})] \\ &= \sum_{k=1}^M \mathbb{E} \int_{T-(k-1)h}^{T-kh} \left(-\partial_t u(T-t, \bar{X}_t) + \partial_x u(T-t, \bar{X}_t) \bar{b}(\bar{X}_{\eta(t)}, U_{\eta(t)}) + \frac{1}{2} \partial_{xx}^2 u(T-t, \bar{X}_t) \right) dt. \end{aligned}$$

From now on, let us denote $t_k := kh$. Since u satisfies the PDE (4.1), we see that for all $x \in \mathbb{R}^d$ we have $-\partial_t u(T-t, x) + \frac{1}{2} \partial_{xx}^2 u(T-t, x) = -\partial_x u(T-t, x) b(x)$. Hence

$$\begin{aligned} \mathbb{E}[g(\bar{X}_T)] - \mathbb{E}[g(Y_T)] &= \sum_{k=1}^M \mathbb{E} \int_{T-t_{k-1}}^{T-t_k} \partial_x u(T-t, \bar{X}_t) (\bar{b}(\bar{X}_{\eta(t)}, U_{\eta(t)}) - b(\bar{X}_t)) dt \\ &= \sum_{k=1}^M \mathbb{E} \int_{T-t_{k-1}}^{T-t_k} [(\partial_x u(T-t, \bar{X}_t) - \partial_x u(T-t, \bar{X}_{\eta(t)})) (\bar{b}(\bar{X}_{\eta(t)}, U_{\eta(t)}) - b(\bar{X}_t))] dt \\ &\quad + \sum_{k=1}^M \mathbb{E} \int_{T-t_{k-1}}^{T-t_k} [\partial_x u(T-t, \bar{X}_{\eta(t)}) (\bar{b}(\bar{X}_{\eta(t)}, U_{\eta(t)}) - b(\bar{X}_t))] dt \\ &= \sum_{k=1}^M \mathbb{E} \int_{T-t_{k-1}}^{T-t_k} [(\partial_x u(T-t, \bar{X}_t) - \partial_x u(T-t, \bar{X}_{\eta(t)})) (\bar{b}(\bar{X}_{\eta(t)}, U_{\eta(t)}) - b(\bar{X}_{\eta(t)}))] dt \quad (:= E_1) \\ &\quad + \sum_{k=1}^M \mathbb{E} \int_{T-t_{k-1}}^{T-t_k} [(\partial_x u(T-t, \bar{X}_t) - \partial_x u(T-t, \bar{X}_{\eta(t)})) (b(\bar{X}_{\eta(t)}) - b(\bar{X}_t))] dt \quad (:= E_2) \\ &\quad + \sum_{k=1}^M \mathbb{E} \int_{T-t_{k-1}}^{T-t_k} [\partial_x u(T-t, \bar{X}_{\eta(t)}) (\bar{b}(\bar{X}_{\eta(t)}, U_{\eta(t)}) - b(\bar{X}_t))] dt \quad (:= E_3). \end{aligned}$$

We begin by considering the last term, denoted by E_3 . Observe that

$$\begin{aligned} &\mathbb{E} [\partial_x u(T-t, \bar{X}_{\eta(t)}) (\bar{b}(\bar{X}_{\eta(t)}, U_{\eta(t)}) - b(\bar{X}_t)) | \bar{X}_{\eta(t)}] \\ &= \mathbb{E} [\partial_x u(T-t, \bar{X}_{\eta(t)}) \mathbb{E} [(b(\bar{X}_{\eta(t)}) - b(\bar{X}_t)) | \bar{X}_{\eta(t)}]] \\ &\leq \Gamma_1 e^{-\gamma_1(T-t)} \mathbb{E} [(1 + |\bar{X}_{\eta(t)}|^{p_1}) \cdot |\mathbb{E} [(b(\bar{X}_{\eta(t)}) - b(\bar{X}_t)) | \bar{X}_{\eta(t)}]|] \\ &\leq \Gamma_1 e^{-\gamma_1(T-t)} (2 + 2C_{IEul}^{(2p_1)})^{1/2} (\mathbb{E} |\mathbb{E} [(b(\bar{X}_{\eta(t)}) - b(\bar{X}_t)) | \bar{X}_{\eta(t)}]|^2)^{1/2}, \end{aligned}$$

where we used $\mathbb{E} |\bar{X}_{\eta(t)}|^{2p_1} \leq C_{IEul}^{(2p_1)}$ due to Remark 2.18. Application of the Itô formula and the fact that $b \in C_b^3$ leads to

$$\begin{aligned} \mathbb{E} [(b(\bar{X}_t) - b(\bar{X}_{\eta(t)})) | \bar{X}_{\eta(t)}] &= \mathbb{E} \left[\int_{\eta(t)}^t b'(\bar{X}_s) \bar{b}(\bar{X}_{\eta(s)}, U_{\eta(s)}) + \frac{1}{2} b''(\bar{X}_s) ds \mid \bar{X}_{\eta(t)} \right] \\ &\leq \frac{1}{2} C_{b''} (t - \eta(t)) + C_{b'} \int_{\eta(t)}^t \mathbb{E} [|\bar{b}(\bar{X}_{\eta(s)}, U_{\eta(s)})| | \bar{X}_{\eta(t)}] ds, \end{aligned}$$

where $C_{b'}$ and $C_{b''}$ are bounds on the derivatives of b . By Minkowski's inequality and (2.40)

$$(\mathbb{E} [|\bar{b}(\bar{X}_{\eta(s)}, U_{\eta(s)})| | \bar{X}_{\eta(t)}])^2 \leq \sigma^2 (|\bar{X}_{\eta(s)}|^2 + 1) h^\alpha + 2L^2 |\bar{X}_{\eta(s)}|^2 + 2|b(0)|^2 \leq C (|\bar{X}_{\eta(s)}|^2 + 1),$$

where $C := \max(\sigma^2 + 2L^2, \sigma^2 + 2|b(0)|^2)$, since $h \leq 1$. Consequently

$$E_3 \leq \bar{C} \sum_{k=1}^M \int_{T-t_{k-1}}^{T-t_k} e^{-\gamma_1(T-t)} (t - \eta(t)) dt,$$

with $\bar{C} := \Gamma_1(2 + 2C_{IEul}^{(2p_1)})^{1/2} (\frac{1}{2}C_{b''}^2 + 2C_{b'}^2(C_{IEul} + 1) \max(\sigma^2 + 2L^2, \sigma^2 + 2|b(0)|^2))^{1/2}$. We conclude our estimation of E_3 by observing that, since $|t - \eta(t)| \leq h$, we have

$$\bar{C} \sum_{k=1}^M \int_{T-t_{k-1}}^{T-t_k} e^{-\gamma_1(T-t)} (t - \eta(t)) dt \leq \bar{C} h \sum_{k=1}^M \int_{T-t_{k-1}}^{T-t_k} e^{-\gamma_1(T-t)} dt = \bar{C} h \int_0^T e^{-\gamma_1(T-t)} dt \leq \bar{C} \frac{h}{\gamma_1}.$$

That completes the estimates of the third term E_3 . To estimate the second term E_2 we calculate

$$\begin{aligned} & \sum_{k=1}^M \mathbb{E} \int_{T-t_{k-1}}^{T-t_k} [(\partial_x u(T-t, \bar{X}_t) - \partial_x u(T-t, \bar{X}_{\eta(t)})) (b(\bar{X}_{\eta(t)}) - b(\bar{X}_t))] dt \\ & \leq L \sum_{k=1}^M \mathbb{E} \int_{T-t_{k-1}}^{T-t_k} \left[\int_0^1 |\partial_{xx} u(T-t, \alpha \bar{X}_t + (1-\alpha) \bar{X}_{\eta(t)})| d\alpha |\bar{X}_{\eta(t)} - \bar{X}_t|^2 \right] dt \\ & \leq \sum_{k=1}^M \int_{T-t_{k-1}}^{T-t_k} \Gamma_2 e^{-\gamma_2(T-t)} 2^{p_2-1} \mathbb{E} [(1 + |\bar{X}_t|^{p_2} + |\bar{X}_{\eta(t)}|^{p_2}) |\bar{X}_{\eta(t)} - \bar{X}_t|^2] dt \\ & \leq \sqrt{3} 2^{p_2-1} \sum_{k=1}^M \int_{T-t_{k-1}}^{T-t_k} \Gamma_2 e^{-\gamma_2(T-t)} (\mathbb{E}[(1 + |\bar{X}_t|^{2p_2} + |\bar{X}_{\eta(t)}|^{2p_2})])^{1/2} (\mathbb{E}[|\bar{X}_{\eta(t)} - \bar{X}_t|^4])^{1/2} dt. \end{aligned}$$

The first expectation in the integrand can be bounded using Remark 2.18, while for the second we write

$$\begin{aligned} \mathbb{E} |\bar{X}_t - \bar{X}_{\eta(t)}|^4 &= \mathbb{E} \left| \int_{\eta(t)}^t \bar{b}(\bar{X}_{\eta(s)}, U_{\eta(s)}) ds + W(t) - W(\eta(t)) \right|^4 \\ &\leq 8 \left((t - \eta(t))^3 \int_{\eta(t)}^t \mathbb{E} |\bar{b}(\bar{X}_{\eta(s)}, U_{\eta(s)})|^4 ds + 3(t - \eta(t))^2 \right), \end{aligned}$$

where we use $(a + b)^4 \leq 8a^4 + 8b^4$, $\mathbb{E}|W(t) - W(\eta(t))|^4 = 3|t - \eta(t)|^2$ and $\left(\int_a^b |g|\right)^p \leq (b - a)^{p-1} \int_a^b |g|^p$ due to the Hölder inequality.

Since we assume that for all $x \in \mathbb{R}$ we have $\mathbb{E}|\bar{b}(x, U)|^4 \leq C_b^4(1 + |x|^4)$, we can now bound $|t - \eta(t)| \leq h$ and complete the estimate for E_2 in a similar way as for E_3 .

It now remains to deal with E_1 . We have

$$\begin{aligned} & \sum_{k=1}^M \mathbb{E} \int_{T-t_{k-1}}^{T-t_k} [(\partial_x u(T-t, \bar{X}_t) - \partial_x u(T-t, \bar{X}_{\eta(t)})) (\bar{b}(\bar{X}_{\eta(t)}, U_{\eta(t)}) - b(\bar{X}_{\eta(t)}))] dt \\ & \leq \sum_{k=1}^M \int_{T-t_{k-1}}^{T-t_k} (\mathbb{E} |\partial_x u(T-t, \bar{X}_t) - \partial_x u(T-t, \bar{X}_{\eta(t)})|^2)^{1/2} (\mathbb{E} |\bar{b}(\bar{X}_{\eta(t)}, U_{\eta(t)}) - b(\bar{X}_{\eta(t)})|^2)^{1/2} dt \\ & \leq \sigma(1 + C_{IEul})^{1/2} \sum_{k=1}^M \int_{T-t_{k-1}}^{T-t_k} (\mathbb{E} |\partial_x u(T-t, \bar{X}_t) - \partial_x u(T-t, \bar{X}_{\eta(t)})|^2)^{1/2} dt, \end{aligned}$$

where we used (2.39) and $h \leq 1$. For the remaining expectation, since $u \in C^{1,3}$, we have

$$\begin{aligned} & \partial_x u(T-t, \bar{X}_t) - \partial_x u(T-t, \bar{X}_{\eta(t)}) \\ &= \int_{\eta(t)}^t \left(-\partial_{tx}^2 u(T-s, \bar{X}_s) + \partial_{xx}^2 u(T-s, \bar{X}_s) \bar{b}(\bar{X}_{\eta(s)}, U_{\eta(s)}) + \frac{1}{2} \partial_{xxx}^3 u(T-s, \bar{X}_s) \right) ds. \end{aligned}$$

Since $u \in C^{1,3}([0, T], \mathbb{R})$, we can infer from (4.1) that it satisfies

$$\begin{cases} \partial_{tx} u(t, x) - \partial_{xx}^2 u(t, x) b(x) - \frac{1}{2} \partial_{xxx}^3 u(t, x) = \partial_x u(t, x) b'(x), & (t, x) \in [0, \infty) \times \mathbb{R}, \\ \partial_x u(0, x) = g'(x). \end{cases} \quad (4.3)$$

Hence we have

$$\begin{aligned} & \partial_x u(T-t, \bar{X}_t) - \partial_x u(T-t, \bar{X}_{\eta(t)}) \\ &= \int_{\eta(t)}^t \left(\partial_{xx}^2 u(T-s, \bar{X}_s) (\bar{b}(\bar{X}_{\eta(s)}, U_{\eta(s)}) - b(\bar{X}_s)) - \partial_x u(T-s, \bar{X}_s) b'(\bar{X}_s) \right) ds \\ &\leq \int_{\eta(t)}^t \left(\Gamma_2 e^{-\gamma_2(T-s)} (1 + |\bar{X}_s|^{p_2}) |\bar{b}(\bar{X}_{\eta(s)}, U_{\eta(s)}) - b(\bar{X}_s)| + C_{b'} \Gamma_1 e^{-\gamma_1(T-s)} (1 + |\bar{X}_s|^{p_1}) \right) ds. \end{aligned}$$

Now we can use the Hölder inequality as above with $p = 2$, take the expectation, use the Cauchy-Schwarz inequality for \mathbb{E} to separate the terms involving $|\bar{X}_s|$, $|\bar{b}(\bar{X}_{\eta(s)}, U_{\eta(s)})|$ and $|b(\bar{X}_s)|$, use the linear growth conditions (7.1) and (2.40) for b and \bar{b} , respectively, and then use the uniform bounds for $\mathbb{E}|\bar{X}_s|^p$ for all $p \geq 1$ as in the previous step. All these efforts combined give us a constant $\hat{C} > 0$ such that

$$\partial_x u(T-t, \bar{X}_t) - \partial_x u(T-t, \bar{X}_{\eta(t)}) \leq \hat{C}(t - \eta(t)) e^{-(\gamma_1 \wedge \gamma_2)(T-t)}.$$

Proceeding as before, we finish the estimate of E_1 and hence the entire proof. \square

5 Further possible extensions

We would like to highlight the robustness of the general approach presented in this paper by briefly discussing several possible extensions of our results to more general settings. For the lack of space, we do not cover these additional cases in detail. However, they are all based on the idea that whenever for a Markov chain $(X_k)_{k=0}^\infty$ we have a one-step contraction of the form (2.12) in some (pseudo-)distance function f (see [EM19, QH19] and the references therein for further examples of such contractions), we can then use the properties of f to obtain a perturbation inequality such as (2.17) and employ it to study error bounds for sampling algorithms based on $(X_k)_{k=0}^\infty$.

5.1 Non-constant discretisation parameter

Our analysis of ULA in Section 2.3 is based on the one-step analysis of Euler schemes in Section 2.2. However, we do not need to use the same discretisation parameter in each step and the proofs of Theorems 2.9 and 1.2 still work in the same way if the discretisation parameters vary between steps. More precisely, let us consider a decreasing sequence $(h_k)_{k=0}^\infty$ such that $h_0 \leq \min \left(\frac{K}{4L^2}, \frac{4}{K}, \frac{1}{2L}, \frac{2c_0 \ln \frac{3}{2}}{27L^2\mathcal{R}^2}, \frac{\mathcal{R}^2}{4}, \frac{c_0^2(\ln 2)^2}{144L^2\mathcal{R}^2} \right)$, so that all h_k for $k \geq 0$ are within the range required in Theorem 2.9. We consider a Markov chain given by

$$X_{k+1} = X_k + h_k b(X_k) + \sqrt{h_k} Z_{k+1}.$$

Then in the proof of Theorem 2.9 for each $k \geq 0$ we have

$$\mathbb{E}f(|G_{k+1} - Y_{(k+1)h_k}|) \leq (1 - ch_k)\mathbb{E}f(|G_k - Y_{kh_{k-1}}|) + C_{ult}h_k^{3/2}$$

and hence, instead of (2.37), we obtain

$$\mathbb{E}f(|G_{k+1} - Y_{(k+1)h_k}|) \leq \left(\prod_{l=0}^k (1 - ch_l) \right) \mathbb{E}f(|G_0 - Y_0|) + C_{ult} \sum_{j=0}^k \left(\prod_{l=1}^j (1 - ch_{k-(l-1)}) h_{k-j}^{3/2} \right),$$

with the convention $\prod_{l=1}^0 (\dots) = 1$. This leads to

$$\begin{aligned} W_2(\mathcal{L}(X_k), \pi) &\leq \left(A \left(\prod_{l=0}^{k-1} (1 - ch_l) \right) \mathbb{E}f(|X_0 - V_0|) \right)^{1/2} \\ &\quad + \left(AC_{ult} \sum_{j=0}^{k-1} \left(\prod_{l=1}^j (1 - ch_{k-(l-1)}) h_{k-j}^{3/2} \right) \right)^{1/2}, \end{aligned}$$

where $V_0 \sim \pi$, and bounds for W_1 can be obtained in a similar way.

5.2 Bounds in the total variation distance

In order to obtain error bounds for ULA in the total variation distance, we can consider a (pseudo-)distance function with a discontinuity at zero. Namely, if $f : [0, \infty) \rightarrow [0, \infty)$ is of the form

$$f(r) := a\mathbf{1}_{(0, \infty)}(r) + f_1(r), \quad (5.1)$$

where $a > 0$ and $f_1 : [0, \infty) \rightarrow [0, \infty)$ is a continuous, non-decreasing function, then for all probability measures μ and ν on \mathbb{R}^d we have $\|\mu - \nu\|_{TV} \leq 2a^{-1}W_f(\mu, \nu)$ (recall that $\|\mu - \nu\|_{TV} = 2 \inf_{\pi \in \Pi(\mu, \nu)} \int \mathbf{1}_{(0, \infty)}(|x - y|) \pi(dx dy)$ and note that $W_{\mathbf{1}_{(0, \infty)}}(\mu, \nu) = \frac{1}{2}\|\mu - \nu\|_{TV}$). Hence, if for a Markov chain $(X_k)_{k=0}^\infty$ we obtain a one-step contraction such as (2.12) in a (pseudo-)distance based on the function f given by (5.1), then we can use the same reasoning as in Section 2.3 to control $\|\mathcal{L}(X_k) - \pi\|_{TV}$. Obtaining such contractions is indeed possible under our Assumptions 1.1, see e.g. Theorem 2.10 in [EM19]. Note that the distance function and all the constants there are explicit (although rather complicated). We leave the details to the interested reader.

5.3 Bounds without requiring contractivity of the drift even at infinity

Another possible extension of our results would be to remove the assumption (1.5) of contractivity at infinity of the drift and replace it with the weaker condition (2.33), i.e., to assume that there exist constants $M_1, M_2 > 0$ such that for all $x \in \mathbb{R}^d$ we have

$$\langle b(x), x \rangle \leq M_2 - M_1|x|^2. \quad (5.2)$$

Note that in Lemma 2.11 we showed that (1.5) implies (5.2) so the latter condition is indeed more general. Note that an extension in this direction in the context of Wasserstein contractions for diffusion processes has been recently obtained in [EGZ19], which generalized the results from [Ebe16]. An analogous result directly at the level of Euler schemes has been obtained in [EM19], see Theorem 6.1 and Example 6.2 therein. Namely, the authors of [EM19] showed that under Assumptions 1.1 with (1.5) replaced by (5.2), one can construct a pseudo-distance ρ_a of the form

$$\rho_a(x, y) := (a + C(V(x) + V(y))) \mathbf{1}_{(0, \infty)}(|x - y|) + f_1(|x - y|) \quad (5.3)$$

and a coupling (X', Y') of the Euler scheme (2.1) with a transition kernel p , such that (X', Y') is contractive in W_{ρ_a} , i.e., using the notation from Corollary 2.2,

$$W_{\rho_a}(\mu p, \nu p) \leq (1 - ch)W_{\rho_a}(\mu, \nu) \quad (5.4)$$

with some constant $c > 0$, for sufficiently small $h > 0$ and for all probability measures μ and ν on \mathbb{R}^d . In (5.3), a and C are positive constants, V is a Lyapunov function and f_1 is a continuous non-decreasing function. Since ρ_a , similarly as f in Subsection 5.2, dominates $\mathbf{1}_{(0, \infty)}$ (up to a multiplicative constant) the contraction in (5.4) allows us to control $\|\mathcal{L}(X_k) - \pi\|_{TV}$, where $(X_k)_{k=0}^\infty$ is a scheme with the transition kernel p . Moreover, ρ_a also dominates the truncated Wasserstein distance W_1^* given by

$$W_1^*(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} (|x - y| \wedge 1) \pi(dx dy).$$

Hence (5.4) allows us to control also $W_1^*(\mathcal{L}(X_k), \pi)$. By tracing the proof of our Theorem 2.16, we can then easily extend all these results to chains with inaccurate drifts. Again, for the lack of space, we leave the details to the reader.

6 Appendix: Coupling for Euler schemes

In this section we provide a proof that the random vector (X', Y') presented in (2.7) is indeed a coupling of two copies of X' . Let us start with the case of $H = \infty$, i.e., without the truncation based on the distance of the processes before the jump.

Theorem 6.1. *The joint distribution of the 2d-dimensional random vector (X', Y') defined by (2.7) is a coupling of $N(\hat{x}, hI)$ and $N(\hat{y}, hI)$.*

Proof. We need to show that $Y' \sim N(\hat{y}, hI)$. To this end, we will show that for any continuous bounded function g we have

$$\mathbb{E}g(Y') = \mathbb{E}g(\hat{y} + \sqrt{h}Z),$$

where $Z \sim N(0, I)$. Straight from the definition of Y' we get

$$\mathbb{E}g(Y') = I_1 + I_2 + I_3, \quad (6.1)$$

where

$$I_1 = \int_{\mathbb{R}^d} g(\hat{x} + \sqrt{h}z) \frac{\varphi_{\hat{y}, hI}^m(\hat{x} + \sqrt{h}z) \wedge \varphi_{\hat{x}, hI}^m(\hat{x} + \sqrt{h}z)}{\varphi_{\hat{x}, hI}^m(\hat{x} + \sqrt{h}z)} \mathbf{1}_{\{|\sqrt{h}z| < m\}} \varphi_{0, I}(z) dz,$$

$$I_2 = \int_{\mathbb{R}^d} g(\hat{y} + R_{\hat{x}, \hat{y}} \sqrt{h}z) \left(1 - \frac{\varphi_{\hat{y}, hI}^m(\hat{x} + \sqrt{h}z) \wedge \varphi_{\hat{x}, hI}^m(\hat{x} + \sqrt{h}z)}{\varphi_{\hat{x}, hI}^m(\hat{x} + \sqrt{h}z)} \right) \mathbf{1}_{\{|\sqrt{h}z| < m\}} \varphi_{0, I}(z) dz$$

$$\text{and } I_3 = \int_{\mathbb{R}^d} g(\hat{y} + \sqrt{h}z) \mathbf{1}_{\{|\sqrt{h}z| \geq m\}} \varphi_{0, I}(z) dz.$$

By substituting $u = \hat{x} + \sqrt{h}z$ and using the fact that $\frac{1}{\sqrt{h}}\varphi_{0, I}(\frac{u-\hat{x}}{\sqrt{h}}) = \varphi_{\hat{x}, hI}(u)$, we see that

$$I_1 = \int_{\mathbb{R}^d} g(u) (\varphi_{\hat{y}, hI}^m(u) \wedge \varphi_{\hat{x}, hI}^m(u)) du.$$

In a similar way, we can substitute $u = \hat{y} + \sqrt{h}z$ to see that

$$I_3 = \int_{\mathbb{R}^d} g(u) \mathbf{1}_{\{|u-\hat{y}| \geq m\}} \varphi_{\hat{y},hI}(u) du.$$

In order to deal with I_2 , we substitute $u = \hat{y} + R_{\hat{x},\hat{y}}\sqrt{h}z$ and notice that then $\sqrt{h}z = R_{\hat{x},\hat{y}}(u - \hat{y})$ since $R_{\hat{x},\hat{y}}$ is an involution. Moreover, since $R_{\hat{x},\hat{y}}$ is an isometry and $|\det R_{\hat{x},\hat{y}}| = 1$, we obtain

$$I_2 = \int_{\mathbb{R}^d} g(u) \left(1 - \frac{\varphi_{\hat{y},hI}^m(\hat{x} + R_{\hat{x},\hat{y}}(u - \hat{y})) \wedge \varphi_{\hat{x},hI}^m(\hat{x} + R_{\hat{x},\hat{y}}(u - \hat{y}))}{\varphi_{\hat{x},hI}^m(\hat{x} + R_{\hat{x},\hat{y}}(u - \hat{y}))} \right) \mathbf{1}_{\{|u-\hat{y}| < m\}} \varphi_{\hat{y},hI}(u) du.$$

Here we also used $\frac{1}{\sqrt{h}}\varphi_{0,I}\left(\frac{R_{\hat{x},\hat{y}}(u-\hat{y})}{\sqrt{h}}\right) = \varphi_{\hat{y},hI}(u)$, which again follows from $R_{\hat{x},\hat{y}}$ being an isometry. If we now note that $R_{\hat{x},\hat{y}}(\hat{y} - \hat{x}) = \hat{x} - \hat{y}$ and hence

$$\varphi_{\hat{y},hI}^m(\hat{x} + R_{\hat{x},\hat{y}}(u - \hat{y})) = \varphi_{\hat{y}-\hat{x},hI}^m(R_{\hat{x},\hat{y}}(u - \hat{y})) = \varphi_{\hat{x},hI}^m(u),$$

we see that

$$I_2 = \int_{\mathbb{R}^d} g(u) \varphi_{\hat{y},hI}^m(u) du - \int_{\mathbb{R}^d} g(u) (\varphi_{\hat{x},hI}^m(u) \wedge \varphi_{\hat{y},hI}^m(u)) du,$$

which implies that $I_1 + I_2 + I_3 = \int_{\mathbb{R}^d} g(u) \varphi_{\hat{y},hI}(u) du$. \square

Now we can consider the general case of (X', Y') given by (2.7) with $H \in (0, \infty)$. We can easily check that this is indeed a coupling of $N(\hat{x}, hI)$ and $N(\hat{y}, hI)$ by applying Theorem 6.1 and observing that for any continuous bounded function g we have

$$\mathbb{E}g(Y') = \mathbf{1}_{\{\hat{r} > H\}} \mathbb{E}g(\hat{y} + \sqrt{h}Z) + \mathbf{1}_{\{\hat{r} \leq H\}} (I_1 + I_2 + I_3),$$

where I_1 , I_2 and I_3 denote the expressions that appear on the right hand side of (6.1). Since we know from the proof of Theorem 6.1 that $I_1 + I_2 + I_3 = \int_{\mathbb{R}^d} g(u) \varphi_{\hat{y},hI}(u) du$, we get our assertion.

7 Appendix: Proofs of lemmas from Section 2

7.1 Proof of Lemma 2.11

Proof. If $|x| > \max\left(\mathcal{R}, \frac{2|b(0)|}{K}\right)$ then

$$\langle b(x), x \rangle = \langle b(x) - b(0), x \rangle + \langle b(0), x \rangle \leq -K|x|^2 + |b(0)||x| \leq -K|x|^2 + \frac{K}{2|b(0)|}|b(0)||x|^2 = -\frac{K}{2}|x|^2.$$

On the other hand, if $|x| \leq \max\left(\mathcal{R}, \frac{2|b(0)|}{K}\right)$, then

$$\langle b(x), x \rangle \leq |b(x) - b(0)||x| + |b(0)||x| \leq L \left(\max\left(\mathcal{R}, \frac{2|b(0)|}{K}\right) \right)^2 + |b(0)| \max\left(\mathcal{R}, \frac{2|b(0)|}{K}\right).$$

\square

7.2 Proof of Lemma 2.12

Proof. By the Itô formula combined with (2.33), for $\tau_n := \inf\{t > 0 : |Y_t| > n\}$ we have

$$\begin{aligned}\mathbb{E}|Y_{t \wedge \tau_n}|^2 &= \mathbb{E}|Y_0|^2 + 2\mathbb{E} \int_0^{t \wedge \tau_n} \langle b(Y_s), Y_s \rangle ds + \mathbb{E}(t \wedge \tau_n)d \\ &\leq \mathbb{E}|Y_0|^2 + t(2M_2 + d) - 2M_1 \int_0^t \mathbb{E}|Y_{s \wedge \tau_n}|^2 ds.\end{aligned}$$

By the Gronwall inequality and the Fatou lemma, we get $\mathbb{E}|Y_t|^2 \leq (\mathbb{E}|Y_0|^2 + t(2M_2 + d)) e^{-2M_1 t}$. Note that the function $t \mapsto te^{-2M_1 t}$ is bounded by $\frac{1}{2M_1}$ for all $t \geq 0$, which implies the desired bound. \square

7.3 Proof of Lemma 2.13

Proof. Straight from the definition (2.25) of $(X_k)_{k=0}^\infty$ we have

$$\mathbb{E}|X_{k+1}|^2 = \mathbb{E}|X_k|^2 + \mathbb{E}|hb(X_k) + \sqrt{h}Z_{k+1}|^2 + 2\mathbb{E}\langle X_k, hb(X_k) + \sqrt{h}Z_{k+1} \rangle.$$

Now note that (1.4) implies that for all $x \in \mathbb{R}^d$ we have

$$|b(x)|^2 \leq 2L^2|x|^2 + 2|b(0)|^2. \quad (7.1)$$

This, together with the fact that $\mathbb{E}|Z_{k+1}|^2 = d$ and (2.33), implies

$$\mathbb{E}|X_{k+1}|^2 \leq \mathbb{E}|X_k|^2 (1 + 4h^2L^2 - 2M_1h) + (4h^2|b(0)|^2 + 2hd + 2hM_2).$$

Hence we have

$$\mathbb{E}|X_{k+1}|^2 \leq \mathbb{E}|X_0|^2 (1 + 4h^2L^2 - 2M_1h)^{k+1} + 2h(2h|b(0)|^2 + d + M_2) \sum_{j=0}^k (1 + 4h^2L^2 - 2M_1h)^j.$$

Note that if $|1 + 4h^2L^2 - 2M_1h| < 1$, which is equivalent to $h < \frac{M_1}{2L^2} = \frac{K}{4L^2}$, then we can bound the finite sum on the right hand side by the infinite sum, and the expression next to $\mathbb{E}|X_0|^2$ by 1, for any $k \geq 0$. This completes the proof. \square

7.4 Proof of Lemma 2.14

Proof. We have

$$\begin{aligned}\mathbb{E} \left| \int_{kh}^{(k+1)h} b(Y_s) ds - hb(Y_{kh}) \right|^2 &= \mathbb{E} \left| \int_{kh}^{(k+1)h} (b(Y_s) - b(Y_{kh})) ds \right|^2 \leq h \mathbb{E} \int_{kh}^{(k+1)h} |b(Y_s) - b(Y_{kh})|^2 ds \\ &\leq L^2 h \mathbb{E} \int_{kh}^{(k+1)h} |Y_s - Y_{kh}|^2 ds \leq 2L^2 h \mathbb{E} \int_{kh}^{(k+1)h} \left(\left| \int_{kh}^s b(Y_r) dr \right|^2 + |W_s - W_{kh}|^2 \right) ds \\ &\leq 2L^2 h \mathbb{E} \int_{kh}^{(k+1)h} \left((s - kh) \int_{kh}^s |b(Y_r)|^2 dr + |W_s - W_{kh}|^2 \right) ds \\ &\leq 2L^2 h \left(\left(\sup_{r \leq (k+1)h} \mathbb{E}|b(Y_r)|^2 \right) \int_{kh}^{(k+1)h} (s - kh)^2 ds + d \int_{kh}^{(k+1)h} (s - kh) ds \right) \\ &= 2L^2 h \left(\left(\sup_{r \leq (k+1)h} \mathbb{E}|b(Y_r)|^2 \right) \frac{h^3}{3} + d \frac{h^2}{2} \right) = L^2 h^3 \left(\left(\sup_{r \leq (k+1)h} \mathbb{E}|b(Y_r)|^2 \right) \frac{2h}{3} + d \right).\end{aligned}$$

Now we combine (7.1) with Lemma 2.12 and we obtain the desired bound. \square

7.5 Proof of Lemma 2.17

Proof. We argue similarly as in the proof of Lemma 2.13. From the definition (1.10) of $(\bar{X}_k)_{k=0}^\infty$ we have

$$\mathbb{E}|\bar{X}_{k+1}|^2 = \mathbb{E}|\bar{X}_k|^2 + \mathbb{E}|h\bar{b}(\bar{X}_k, U_k) + \sqrt{h}Z_{k+1}|^2 + 2\mathbb{E}\langle \bar{X}_k, h\bar{b}(\bar{X}_k, U_k) + \sqrt{h}Z_{k+1} \rangle.$$

Note that $\mathbb{E}\mathbb{E}[\langle \bar{X}_k, h\bar{b}(\bar{X}_k, U_k) + \sqrt{h}Z_{k+1} \rangle | \bar{X}_k] = \mathbb{E}\langle \bar{X}_k, hb(\bar{X}_k) \rangle$. Furthermore, by (2.40) we have

$$\mathbb{E}[|\bar{b}(\bar{X}_k, U_k)|^2 | \bar{X}_k] \leq \sigma^2(|\bar{X}_k|^2 + 1)h^\alpha + 2L^2|\bar{X}_k|^2 + 2|b(0)|^2.$$

Hence, using $\mathbb{E}|Z_{k+1}|^2 = d$ and (2.33), we get

$$\mathbb{E}|\bar{X}_{k+1}|^2 \leq \mathbb{E}|\bar{X}_k|^2 (1 + 2\sigma^2h^{2+\alpha} + 4h^2L^2 - 2M_1h) + (2\sigma^2h^{2+\alpha} + 4h^2|b(0)|^2 + 2hd + 2hM_2).$$

Thus we have

$$\begin{aligned} \mathbb{E}|\bar{X}_{k+1}|^2 &\leq \mathbb{E}|\bar{X}_0|^2 (1 + 2\sigma^2h^2 + 4h^2L^2 - 2M_1h)^{k+1} \\ &\quad + (2\sigma^2h^2 + 4h^2|b(0)|^2 + 2hd + 2hM_2) \sum_{j=0}^k (1 + 2\sigma^2h^2 + 4h^2L^2 - 2M_1h)^j, \end{aligned}$$

where we used $h^{2+\alpha} \leq h^2$ for $h \leq 1$. Note that our upper bound on h guarantees that $|1 + 2\sigma^2h^2 + 4h^2L^2 - 2M_1h| < 1$ and hence we can bound the finite sum on the right hand side by the infinite sum. Moreover, we bound the expression next to $\mathbb{E}|X_0|^2$ by 1, which finishes the proof. \square

8 Appendix: Proofs of lemmas from Section 3

The proofs in this section are based on calculations from Section 6 in [EM19].

8.1 Proof of Lemma 3.1

Proof. It is sufficient to consider the case of $\hat{r} \leq H$ since for $\hat{r} > H$ we apply the synchronous coupling. Observe that $R' = |X' - Y'|$ can take values 0 (when the coupled processes jump to the same point), \hat{r} (when they move synchronously) or

$$\left| \hat{x} - \hat{y} + 2 \frac{\hat{x} - \hat{y}(\hat{x} - \hat{y})^T}{|\hat{x} - \hat{y}|^2} \sqrt{h}Z \right|,$$

where $Z \sim N(0, I)$, when they are reflected. Hence we see that in the reflection case $X' - Y'$ is a sum of two parallel vectors and we have $R' = \hat{r} + 2\sqrt{h}\langle \frac{\hat{x} - \hat{y}}{|\hat{x} - \hat{y}|}, Z \rangle$. However, $\langle \frac{\hat{x} - \hat{y}}{|\hat{x} - \hat{y}|}, Z \rangle$ can be interpreted as a one-dimensional projection of a Gaussian random vector and hence without loss of generality we can assume that $R' - \hat{r} = 2\sqrt{h}W$, where $W \sim N(0, 1)$ is a one-dimensional Gaussian random variable.

From the definition of our coupling, we see that if $\hat{r} \leq H$, then we have

$$\begin{aligned} \mathbb{E}[R' | U] - \hat{r} &= 2 \int_{-\infty}^0 |t| \left(1 - \frac{\varphi_{\hat{r},h}^m(t)}{\varphi_{0,h}^m(t)} \right) \mathbf{1}_{\{|t| \leq m\}} \varphi_{0,h}(t) dt - 2 \int_0^{\hat{r}/2} |t| \left(1 - \frac{\varphi_{\hat{r},h}^m(t)}{\varphi_{0,h}^m(t)} \right) \mathbf{1}_{\{|t| \leq m\}} \varphi_{0,h}(t) dt \\ &\quad + \int_{-\infty}^{\hat{r}/2} (-\hat{r}) \frac{\varphi_{\hat{r},h}^m(t)}{\varphi_{0,h}^m(t)} \mathbf{1}_{\{|t| \leq m\}} \varphi_{0,h}(t) dt + \int_{\hat{r}/2}^{\infty} (-\hat{r}) \mathbf{1}_{\{|t| \leq m\}} \varphi_{0,h}(t) dt, \end{aligned}$$

where the first two terms correspond to the reflection and the next two terms correspond to the situation in which the coupled processes jump to the same point. By making the substitution $u = t - \hat{r}$ in the second and the fourth integrals we get

$$\begin{aligned}\mathbb{E}[R' | U] - \hat{r} &= \int_{-\infty}^0 (-2t) \varphi_{0,h}^m(t) dt + \int_{-\infty}^{-\hat{r}} (2u + 2\hat{r}) \varphi_{0,h}^m(u) du + \int_0^{\hat{r}/2} (-2t) \varphi_{\hat{r},h}^m(t) dt \\ &\quad + \int_{-\hat{r}}^{-\hat{r}/2} (2u + 2\hat{r}) \varphi_{\hat{r},h}^m(u) du + 2 \int_{\hat{r}/2}^{\infty} (-\hat{r}) \varphi_{0,h}^m(t) dt \\ &= \int_{-\infty}^{\hat{r}/2} (-2t) \varphi_{\hat{r},h}^m(t) dt + \int_{-\infty}^{-\hat{r}/2} 2\hat{r} \varphi_{\hat{r},h}^m(u) du + \int_{\hat{r}/2}^{\infty} (-2\hat{r}) \varphi_{0,h}^m(t) dt \\ &\quad + \int_{-\infty}^{-\hat{r}/2} 2u \varphi_{0,h}^m(u) du = \int_{-\hat{r}/2}^{\hat{r}/2} (-2t) \varphi_{0,h}^m(t) dt = 0,\end{aligned}$$

which finishes the proof. \square

Remark 8.1. Note that if instead of the coupling specified in (2.23), we would just use the coupling in which the processes are always reflected, we would get

$$\mathbb{E}[R' | U] - \hat{r} = 2 \int_{-\infty}^0 |t| \varphi_{0,h}(t) dt - 2 \int_0^{\hat{r}/2} |t| \varphi_{0,h}(t) dt + 2 \int_{\hat{r}}^{\infty} |t| \varphi_{0,h}(t) dt,$$

which clearly is positive for all \hat{r} , and hence the assertion of Lemma 3.1 fails.

8.2 Proof of Lemma 3.2

Proof. First observe that just like in Lemma 3.1, without loss of generality we can assume that $R' - \hat{r} = 2\sqrt{h}W$, where $W \sim N(0, 1)$. We have

$$\mathbb{E}[(R' - \hat{r})^2 \mathbf{1}_{\{R' \in I_{\hat{r}}\}} | U] \geq \int_{-\infty}^{\hat{r}/2} (-2t)^2 \left(1 - \frac{\varphi_{\hat{r},h}^m(t)}{\varphi_{0,h}^m(t)}\right) \varphi_{0,h}(t) \mathbf{1}_{\{|t| \leq m\}} \mathbf{1}_{\{\hat{r} - 2t \in I_{\hat{r}}\}} \mathbf{1}_{\{|\hat{r}| \leq H\}} dt.$$

We can therefore skip the condition $|\hat{r}| \leq H$ from now on in our calculations and just add it again at the very end. Let us deal first with the case of $\hat{r} \geq \sqrt{h}$. Then $I_{\hat{r}} = (\hat{r} - \sqrt{h}, \hat{r})$ and we see that $\hat{r} - 2t \in (\hat{r} - \sqrt{h}, \hat{r})$ if and only if $t \in (0, \sqrt{h}/2)$. Hence the integral above is equal to

$$4 \int_0^{\sqrt{h}/2 \wedge m} t^2 \left(1 - \frac{\varphi_{\hat{r},h}^m(t)}{\varphi_{0,h}^m(t)}\right) \varphi_{0,h}(t) dt. \quad (8.1)$$

Now observe that

$$\frac{\varphi_{\hat{r},h}^m(t)}{\varphi_{0,h}^m(t)} = e^{-\frac{|t-\hat{r}|^2}{2h}} \mathbf{1}_{\{|t-\hat{r}| \leq m\}} e^{\frac{|t|^2}{2h}} \leq e^{-\frac{|t-\hat{r}|^2}{2h}} e^{\frac{|t|^2}{2h}} = e^{-\frac{1}{2h}(t^2 - 2t\hat{r} + \hat{r}^2 - t^2)} = e^{-\frac{1}{h}(\frac{\hat{r}^2}{2} - t\hat{r})}.$$

Combining this with $h \leq 4m^2$, we see that the integral in (8.1) is bounded from below by

$$4 \int_0^{\sqrt{h}/2} t^2 \left(1 - e^{-\frac{1}{h}(\frac{\hat{r}^2}{2} - t\hat{r})}\right) \varphi_{0,h}(t) dt.$$

Now we make the substitution $u = \frac{1}{\sqrt{h}}t$ which gives

$$4 \int_0^{1/2} \sqrt{h} h u^2 \left(1 - e^{-\frac{1}{h}(\frac{\hat{r}^2}{2} - \sqrt{h} u \hat{r})}\right) \frac{1}{\sqrt{2\pi h}} e^{-\frac{h u^2}{2h}} du = 4h \int_0^{1/2} u^2 \left(1 - e^{-\frac{\hat{r}}{\sqrt{h}}(u - \frac{\hat{r}}{2\sqrt{h}})}\right) \varphi_{0,1}(u) du. \quad (8.2)$$

Now observe that the function $s \mapsto s(u - s/2)$ is decreasing for $s \geq u$. Moreover, we have $\hat{r}/\sqrt{h} \geq 1 \geq u$ and thus we can bound the integral in (8.2) from below by

$$4h \int_0^{1/2} u^2 \left(1 - e^{u-\frac{1}{2}}\right) \varphi_{0,1}(u) du,$$

which finishes the proof for the case of $\hat{r} \geq \sqrt{h}$.

Assume now that $\hat{r} \leq \sqrt{h}$. Then $I_{\hat{r}} = (0, \hat{r} + \sqrt{h})$ and $\hat{r} - 2t \in I_{\hat{r}}$ if and only if $t \in (-\frac{\sqrt{h}}{2}, \frac{\hat{r}}{2})$. Hence we have

$$\mathbb{E}[(R' - \hat{r})^2 \mathbf{1}_{\{R' \in I_{\hat{r}}\}} | U] \geq \int_{-\sqrt{h}/2}^0 (-2t)^2 \left(1 - \frac{\varphi_{\hat{r},h}^m(t)}{\varphi_{0,h}^m(t)}\right) \varphi_{0,h}(t) \mathbf{1}_{\{|\hat{r}| \leq H\}} dt.$$

Again for convenience we now skip the condition $\hat{r} \leq H$. After making the substitution $u = \frac{1}{\sqrt{h}}t$ and using the lower bound from the calculations from the previous case, we see that the integral above is bounded from below by

$$4h \int_{-1/2}^0 u^2 \left(1 - e^{\frac{\hat{r}}{\sqrt{h}}(u - \frac{\hat{r}}{2\sqrt{h}})}\right) \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du.$$

Since $u \in (-1/2, 0)$, we easily see that $\frac{\hat{r}}{\sqrt{h}}(u - \frac{\hat{r}}{2\sqrt{h}}) \leq -\frac{\hat{r}^2}{2h} \leq 0$ and since $\hat{r} \leq \sqrt{h}$, we also have $u - \hat{r}/(2\sqrt{h}) \geq -1$, which implies $\frac{\hat{r}}{\sqrt{h}}(u - \frac{\hat{r}}{2\sqrt{h}}) \geq -\frac{\hat{r}}{\sqrt{h}} \geq -1$. We consider the function $g(s) := e^s - 1 - (1 - e^{-1})s$. We see that $g(-1) = g(0) = 0$ and there exists $s_0 \in (-1, 0)$ such that $g'(s) < 0$ for $s \in (-1, s_0)$ and $g'(s) > 0$ for $s \in (s_0, 0)$. This implies that $e^s - 1 \leq (1 - e^{-1})s$ for all $s \in [-1, 0]$ hence our integral is bounded from below by

$$\begin{aligned} & 4h \int_{-1/2}^0 u^2 \left(-\frac{\hat{r}}{\sqrt{h}}(u - \frac{\hat{r}}{2\sqrt{h}})\right) (1 - e^{-1}) \varphi_{0,1}(u) du \\ &= \frac{4h\hat{r}}{\sqrt{h}} (1 - e^{-1}) \int_{-1/2}^0 \left(-u^3 + \frac{\hat{r}}{2\sqrt{h}} u^2\right) \varphi_{0,1}(u) du \geq 4\sqrt{h}\hat{r}(1 - e^{-1}) \int_0^{1/2} u^3 \varphi_{0,1}(u) du. \end{aligned}$$

□

Remark 8.2. Note that in the proof we only use the reflection behaviour and we disregard the possibility of jumping to the same point. This is due to the fact that the probability of jumping to the same point decays exponentially fast with h going to zero and hence it would not contribute to our estimates in a significant way. Hence, for Lemma 3.2 to work, it would be sufficient to take the reflection coupling. However, then our calculations for the first moment in Lemma 3.1 would fail (cf. Remark 8.1). On the other hand, for the synchronous coupling Lemma 3.1 holds while Lemma 3.2 clearly fails. Hence the (truncated) mirror coupling given by (2.7) is the only one for which both these Lemmas work.

Acknowledgement

We would like to thank the associate editor and the anonymous referee for numerous useful suggestions. MBM and AM are supported by the EPSRC grant EP/P003818/1. The majority of this work was completed while MBM and AM were affiliated to King's College London.

References

- [ACB17] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 214–223, International Convention Centre, Sydney, Australia, 06–11 Aug 2017. PMLR.
- [BFFN19] Jack Baker, Paul Fearnhead, Emily B. Fox, and Christopher Nemeth. Control variates for stochastic gradient MCMC. *Stat. Comput.*, 29(3):599–615, 2019.
- [CB18] Xiang Cheng and Peter Bartlett. Convergence of Langevin MCMC in KL-divergence. In Firdaus Janoos, Mehryar Mohri, and Karthik Sridharan, editors, *Proceedings of ALT2018*, volume 83 of *Proceedings of Machine Learning Research*, pages 186–211. PMLR, 2018.
- [CCA⁺18] Xiang Cheng, Niladri S. Chatterji, Yasin Abbasi-Yadkori, Peter L. Bartlett, and Michael I. Jordan. Sharp Convergence Rates for Langevin Dynamics in the Non-convex Setting. *ArXiv e-prints*, May 2018.
- [CFM⁺18] Niladri Chatterji, Nicolas Flammarion, Yian Ma, Peter Bartlett, and Michael Jordan. On the theory of variance reduction for stochastic gradient Monte Carlo. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 764–773, Stockholmsmässan, Stockholm Sweden, 10–15 Jul 2018. PMLR.
- [Dal17] Arnak S. Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 79(3):651–676, 2017.
- [DGW04] H. Djellout, A. Guillin, and L. Wu. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Probab.*, 32(3B):2702–2732, 2004.
- [DK19] Arnak S. Dalalyan and Avetik Karagulyan. User-friendly guarantees for the Langevin Monte Carlo with inaccurate gradient. *Stochastic Process. Appl.*, 2019.
- [DM16] Alain Durmus and Eric Moulines. High-dimensional Bayesian inference via the Unadjusted Langevin Algorithm. *arXiv preprint, arXiv:1605.01559*, 2016.
- [DM17] Alain Durmus and Éric Moulines. Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *Ann. Appl. Probab.*, 27(3):1551–1587, 2017.
- [DRVZ17] Alain Durmus, Gareth O. Roberts, Gilles Vilmart, and Konstantinos C. Zygalakis. Fast Langevin based algorithm for MCMC in high dimensions. *Ann. Appl. Probab.*, 27(4):2195–2237, 2017.
- [Ebe11] Andreas Eberle. Reflection coupling and Wasserstein contractivity without convexity. *C. R. Math. Acad. Sci. Paris*, 349(19-20):1101–1104, 2011.
- [Ebe16] Andreas Eberle. Reflection couplings and contraction rates for diffusions. *Probab. Theory Related Fields*, 166(3-4):851–886, 2016.

- [EGZ19] Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. Quantitative Harris-type theorems for diffusions and McKean-Vlasov processes. *Trans. Amer. Math. Soc.*, 371(10):7135–7173, 2019.
- [EM19] Andreas Eberle and Mateusz B. Majka. Quantitative contraction rates for Markov chains on general state spaces. *Electron. J. Probab.*, 24:Paper No. 26, 36, 2019.
- [FG19] Wei Fang and Michael B. Giles. Multilevel Monte Carlo method for ergodic SDEs without contractivity. *J. Math. Anal. Appl.*, 476(1):149–176, 2019.
- [GDVM16] Jack Gorham, Andrew B. Duncan, Sebastian J. Vollmer, and Lester Mackey. Measuring sample quality with diffusions. *arXiv preprint, arXiv:1611.06972*, 2016.
- [Gil08] Michael B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.
- [Gil15] Michael B. Giles. Multilevel Monte Carlo methods. *Acta Numer.*, 24:259–328, 2015.
- [GMS⁺19] Michael B. Giles, Mateusz B. Majka, Lukasz Szpruch, Sebastian J. Vollmer, and Konstantinos C. Zygalakis. Multi-level Monte Carlo methods for the approximation of invariant measures of stochastic differential equations. *Stat. Comput.*, 2019.
- [Hei01] Stefan Heinrich. Multilevel Monte Carlo methods. In *Large-scale scientific computing*, pages 58–67. Springer, 2001.
- [HZ17] Jonathan H. Huggins and James Zou. Quantifying the accuracy of approximate diffusions and Markov chains. In Aarti Singh and Jerry Zhu, editors, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 382–391, Fort Lauderdale, FL, USA, 20–22 Apr 2017. PMLR.
- [JN09] A. Jentzen and A. Neuenkirch. A random Euler scheme for Carathéodory differential equations. *J. Comput. Appl. Math.*, 224(1):346–359, 2009.
- [JO10] Aldéric Joulin and Yann Ollivier. Curvature, concentration and error estimates for Markov chain Monte Carlo. *Ann. Probab.*, 38(6):2418–2442, 2010.
- [Keb05] Ahmed Kebaier. Statistical Romberg extrapolation: a new variance reduction method and applications to option pricing. *Ann. Appl. Probab.*, 15(4):2681–2705, 2005.
- [Kha12] Rafail Khasminskii. *Stochastic stability of differential equations*, volume 66 of *Stochastic Modelling and Applied Probability*. Springer, Heidelberg, second edition, 2012. With contributions by G. N. Milstein and M. B. Nevelson.
- [Kry99] N. V. Krylov. On Kolmogorov’s equations for finite-dimensional diffusions. In *Stochastic PDE’s and Kolmogorov equations in infinite dimensions (Cetraro, 1998)*, volume 1715 of *Lecture Notes in Math.*, pages 1–63. Springer, Berlin, 1999.
- [KW17] Raphael Kruse and Yue Wu. Error analysis of randomized Runge-Kutta methods for differential equations with time-irregular coefficients. *Comput. Methods Appl. Math.*, 17(3):479–498, 2017.

- [Lin92] Torgny Lindvall. *Lectures on the coupling method*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1992. A Wiley-Interscience Publication.
- [LP02] Damien Lamberton and Gilles Pages. Recursive computation of the invariant distribution of a diffusion. *Bernoulli*, 8(3):367–405, 2002.
- [LW16] Dejun Luo and Jian Wang. Exponential convergence in L^p -Wasserstein distance for diffusion processes without uniformly dissipative drift. *Math. Nachr.*, 289(14-15):1909–1926, 2016.
- [McC99] Robert J. McCann. Exact solutions to the transportation problem on the line. *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.*, 455(1984):1341–1380, 1999.
- [MT04] Grigori Noah Milstein and Michael V. Tretyakov. *Stochastic numerics for mathematical physics*. Scientific Computation. Springer-Verlag, Berlin, 2004.
- [NDH⁺17] Tigran Nagapetyan, Andrew B. Duncan, Leonard Hasenclever, Sebastian J. Vollmer, Lukasz Szpruch, and Konstantinos Zygalakis. The true cost of stochastic gradient Langevin dynamics. *arXiv preprint, arXiv:1706.02692*, 2017.
- [NP19] N. Nüsken and G. Pavliotis. Constructing sampling schemes via coupling: Markov semigroups and optimal transport. *SIAM/ASA Journal on Uncertainty Quantification*, 7(1):324–382, 2019.
- [Pan08] Fabien Panloup. Recursive computation of the invariant measure of a stochastic differential equation driven by a Lévy process. *Ann. Appl. Probab.*, 18(2):379–426, 2008.
- [PM14] Pawel Przybyłowicz and Pawel Morkisz. Strong approximation of solutions of stochastic differential equations with time-irregular coefficients via randomized Euler algorithm. *Appl. Numer. Math.*, 78:80–94, 2014.
- [PP12] Gilles Pagès and Fabien Panloup. Ergodic approximation of the distribution of a stationary diffusion: rate of convergence. *Ann. Appl. Probab.*, 22(3):1059–1100, 2012.
- [PS14] Natesh S. Pillai and Aaron Smith. Ergodicity of Approximate MCMC Chains with Applications to Large Data Sets. *arXiv preprint, arXiv:1405.0182*, May 2014.
- [QH19] Qian Qin and James P. Hobert. Geometric convergence bounds for Markov chains in Wasserstein distance based on generalized drift and contraction conditions. *arXiv e-prints*, page arXiv:1902.02964, Feb 2019.
- [Ric88] John A. Rice. *Mathematical statistics and data analysis*. Wadsworth & Brooks/Cole statistics/probability series. Brooks/Cole Pub. Co., 1988.
- [RRT17] Maxim Raginsky, Alexander Rakhlin, and Matus Telgarsky. Non-convex learning via Stochastic Gradient Langevin Dynamics: a nonasymptotic analysis. In Satyen Kale and Ohad Shamir, editors, *Proceedings of the 2017 Conference on Learning Theory*, volume 65 of *Proceedings of Machine Learning Research*, pages 1674–1703, Amsterdam, Netherlands, 07–10 Jul 2017. PMLR.
- [RS18] Daniel Rudolf and Nikolaus Schweizer. Perturbation theory for Markov chains via Wasserstein distance. *Bernoulli*, 24(4A):2610–2639, 2018.

- [Sha16] Ohad Shamir. Without-replacement sampling for stochastic gradient methods. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 46–54. Curran Associates, Inc., 2016.
- [SZ18] Lukasz Szpruch and Xiling Zhang. V -integrability, asymptotic stability and comparison property of explicit numerical schemes for non-linear SDEs. *Math. Comp.*, 87(310):755–783, 2018.
- [Tal90] Denis Talay. Second-order discretization schemes of stochastic differential systems for the computation of the invariant law. *Stochastics*, 29(1):13–36, 1990.
- [Tal96] M. Talagrand. Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.*, 6(3):587–600, 1996.
- [Tho00] Hermann Thorisson. *Coupling, stationarity, and regeneration*. Probability and its Applications (New York). Springer-Verlag, New York, 2000.
- [TTV16] Yee Whye Teh, Alexandre H. Thiery, and Sebastian J. Vollmer. Consistency and fluctuations for stochastic gradient Langevin dynamics. *J. Mach. Learn. Res.*, 17:Paper No. 7, 33, 2016.
- [Vil09] Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.
- [VZT16] Sebastian J. Vollmer, Konstantinos C. Zygalakis, and Yee Whye Teh. Exploration of the (non-)asymptotic bias and variance of stochastic gradient Langevin dynamics. *J. Mach. Learn. Res.*, 17:Paper No. 159, 45, 2016.
- [WT11] Max Welling and Yee W. Teh. Bayesian learning via stochastic gradient Langevin dynamics. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 681–688, 2011.
- [XSL⁺14] T. Xifara, C. Sherlock, S. Livingstone, S. Byrne, and M. Girolami. Langevin diffusions and the Metropolis-adjusted Langevin algorithm. *Statist. Probab. Lett.*, 91:14–19, 2014.